



PORTFOLIO

STATES BAZAAR

Regional Market Analysis

IGNATIUS EVANS ERLANGGA

What is **STATES BAZAAR ?**

STATES BAZAAR is a fictional company that provides a real transaction data for a specific segment of goods trade from one of the world's leading e-commerce platforms.

The **STATES BAZAAR** dataset was obtained from a website called Kaggle. The purpose of this portfolio is to demonstrate my proficiency in data operations using Microsoft Excel. Through this portfolio, I aim to showcase my ability to **process and analyze data using Excel, as well as my understanding of key business metrics that support data-driven analysis from the dataset.**



Table of Contents



Getting to Know the Data



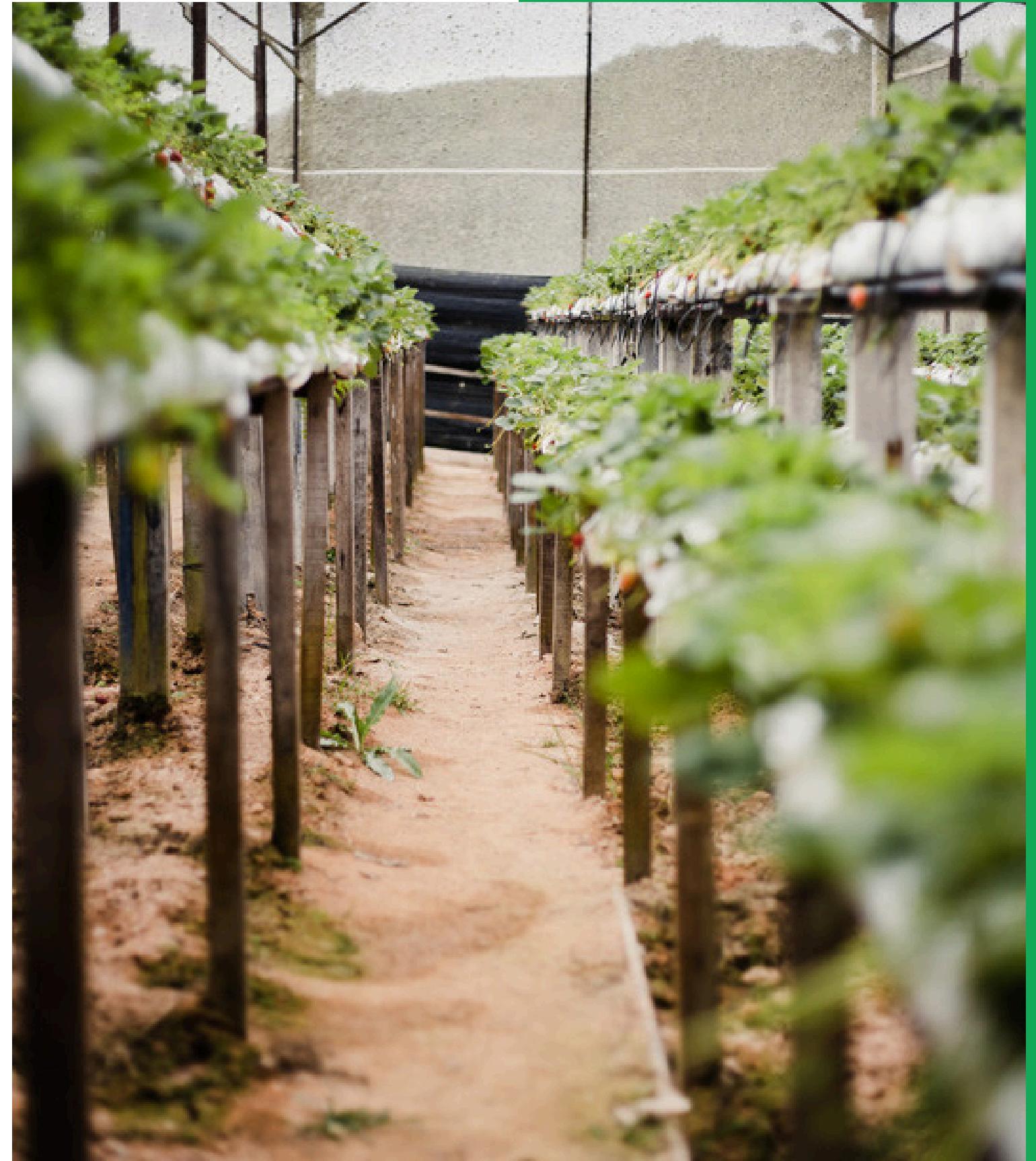
Data Preprocessing



Dashboard



Analysis and Results



What's inside the **STATES BAZAAR's** dataset?

The dataset consists of four sheets, each containing different types of data:

- **Orders:** Contains transaction records
- **Returns:** Contains records of returned (or complained) transactions
- **Manager:** Lists the managers assigned to each region
- **Target:** Contains the sales targets for each region



What's inside the STATES BAZAAR's dataset?

Orders

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	State	Postal Code	Region	Product ID	Category	Sub-Category	Product Name	Sales	Quantity	Discount	Profit	Fixed LOD	Profit Cont.
1	CA-2020-152156	08/11/2020	11/11/2020	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420	South	FUR-BO-1000	Furniture	Bookcases	Bush Somers	261,96	2	0	41,9136	286469,971	0,01%
2	CA-2020-152156	08/11/2020	11/11/2020	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420	South	FUR-CH-1000	Furniture	Chairs	Hon Deluxe F	731,94	3	0	219,582	286469,971	0,08%
3	CA-2020-138688	12/06/2020	16/06/2020	Second Class	DV-13045	Darrin Van H	Corporate	United States	Los Angeles	California	90036	West	OFF-LA-1000	Office Supplies	Labels	Self-Adhesive	14,62	2	0	6,8714	286469,971	0,00%
4	US-2019-108966	11/10/2019	18/10/2019	Standard	Cl-50-20335	Sean O'Donn	Consumer	United States	Fort Lauderdale	Florida	33311	South	FUR-TA-1000	Furniture	Tables	Bretford CR4:	957,5775	5	0,45	-383,031	286469,971	-0,13%
5	US-2019-108966	11/10/2019	18/10/2019	Standard	Cl-50-20335	Sean O'Donn	Consumer	United States	Fort Lauderdale	Florida	33311	South	OFF-ST-1000	Office Supplies	Storage	Eldon Fold 'N	22,368	2	0,2	2,5164	286469,971	0,00%
6	CA-2018-115812	09/06/2018	14/06/2018	Standard	Cl-BH-11710	Brosina Hoff	Consumer	United States	Los Angeles	California	90032	West	FUR-FU-1000	Furniture	Furnishings	Eldon Express	48,86	7	0	14,1694	286469,971	0,00%
7	CA-2018-115812	09/06/2018	14/06/2018	Standard	Cl-BH-11710	Brosina Hoff	Consumer	United States	Los Angeles	California	90032	West	OFF-AR-1000	Office Supplies	Art	Newell 322	7,28	4	0	1,9656	286469,971	0,00%
8	CA-2018-115812	09/06/2018	14/06/2018	Standard	Cl-BH-11710	Brosina Hoff	Consumer	United States	Los Angeles	California	90032	West	TEC-PH-1000	Technology	Phones	Mitel 5320 IP	907,152	6	0,2	90,7152	286469,971	0,03%
9	CA-2018-115812	09/06/2018	14/06/2018	Standard	Cl-BH-11710	Brosina Hoff	Consumer	United States	Los Angeles	California	90032	West	OFF-BI-1000	Office Supplies	Binders	DXL Angle-Vi	18,504	3	0,2	5,7825	286469,971	0,00%
10	CA-2018-115812	09/06/2018	14/06/2018	Standard	Cl-BH-11710	Brosina Hoff	Consumer	United States	Los Angeles	California	90032	West	OFF-AP-1000	Office Supplies	Appliances	Belkin F5C20	114,9	5	0	34,47	286469,971	0,01%
11	CA-2018-115812	09/06/2018	14/06/2018	Standard	Cl-BH-11710	Brosina Hoff	Consumer	United States	Los Angeles	California	90032	West	FUR-TA-1000	Furniture	Tables	Chromcraft R	1706,184	9	0,2	85,3092	286469,971	0,03%
12	CA-2018-115812	09/06/2018	14/06/2018	Standard	Cl-BH-11710	Brosina Hoff	Consumer	United States	Los Angeles	California	90032	West	TEC-PH-1000	Technology	Phones	Konftel 250 C	911,424	4	0,2	68,3568	286469,971	0,02%
13	CA-2021-114412	15/04/2021	20/04/2021	Standard	Cl-AA-10480	Andrew Allen	Consumer	United States	Concord	North Carolina	28027	South	OFF-PA-1000	Office Supplies	Paper	Xerox 1967	15,552	3	0,2	5,4432	286469,971	0,00%
14	CA-2020-161389	05/12/2020	10/12/2020	Standard	Cl-IM-15070	Irene Maddo	Consumer	United States	Seattle	Washington	98103	West	OFF-BI-1000	Office Supplies	Binders	Fellowes PB2	407,976	3	0,2	132,5922	286469,971	0,05%
15	US-2019-118983	22/11/2019	26/11/2019	Standard	Cl-HP-14815	Harold Pawla	Home Office	United States	Fort Worth	Texas	76106	Central	OFF-AP-1000	Office Supplies	Appliances	Holmes Repla	68,81	5	0,8	-123,858	286469,971	-0,04%
16	US-2019-118983	22/11/2019	26/11/2019	Standard	Cl-HP-14815	Harold Pawla	Home Office	United States	Fort Worth	Texas	76106	Central	OFF-BI-1000	Office Supplies	Binders	Storex DuraT	2,544	3	0,8	-3,816	286469,971	0,00%
17	CA-2018-105893	11/11/2018	18/11/2018	Standard	Cl-PK-19075	Pete Kriz	Consumer	United States	Madison	Wisconsin	53711	Central	OFF-ST-1000	Office Supplies	Storage	Stur-D-Stor Sh	665,88	6	0	13,3176	286469,971	0,00%
18	CA-2018-167164	13/05/2018	15/05/2018	Second Class	AG-10270	Alejandro Gru	Consumer	United States	West Jordan	Utah	84084	West	OFF-ST-1000	Office Supplies	Storage	Fellowes Sup	55,5	2	0	9,99	286469,971	0,00%
19	CA-2018-143336	27/08/2018	01/09/2018	Second Class	ZD-21925	Zuschuss Dor	Consumer	United States	San Francisco	California	94109	West	OFF-AR-1000	Office Supplies	Art	Newell 341	8,56	2	0	2,4824	286469,971	0,00%
20	CA-2018-143336	27/08/2018	01/09/2018	Second Class	ZD-21925	Zuschuss Dor	Consumer	United States	San Francisco	California	94109	West	TEC-PH-1000	Technology	Phones	Cisco SPA 50	213,48	3	0,2	16,011	286469,971	0,01%

Column Definitions in the Orders Table:

- Row ID: Index number for each row in the dataset.
- Order ID: A unique identifier assigned to each transaction.
- Order Date: The date when the order was placed.
- Ship Date: The date when the order was delivered to the customer.
- Customer ID: A unique identifier assigned to each customer.
- Segment: The market segment classification of the customer.
- Country: The country of origin of the customer.
- City: The city where the customer is located.
- State: The state or province of the customer's address.
- Postal Code: The postal code of the customer's address.
- Region: The geographical region where the customer is located.
- Product ID: A unique identifier assigned to each product.

What's inside the STATES BAZAAR's dataset?

Orders

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer	Customer Segment	Country	City	State	Postal Code	Region	Product ID	Category	Sub-Category	Product Name	Sales	Quantity	Discount	Profit	Fixed LOD	Profit Cont.	
1	CA-2020-152156	08/11/2020	11/11/2020	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420	South	FUR-BO-1000	Furniture	Bookcases	Bush Somers	261,96	2	0	41,9136	286469,971	0,01%
2	CA-2020-152156	08/11/2020	11/11/2020	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420	South	FUR-CH-1000	Furniture	Chairs	Hon Deluxe F	731,94	3	0	219,582	286469,971	0,08%
3	CA-2020-138688	12/06/2020	16/06/2020	Second Class	DV-13045	Darrin Van H	Corporate	United States	Los Angeles	California	90036	West	OFF-LA-1000	Office Supplies	Labels	Self-Adhesive	14,62	2	0	6,8714	286469,971	0,00%
4	US-2019-108966	11/10/2019	18/10/2019	Standard	Cl:SO-20335	Sean O'Donn	Consumer	United States	Fort Lauderdale	Florida	33311	South	FUR-TA-1000	Furniture	Tables	Bretford CR4:	957,5775	5	0,45	-383,031	286469,971	-0,13%
5	US-2019-108966	11/10/2019	18/10/2019	Standard	Cl:SO-20335	Sean O'Donn	Consumer	United States	Fort Lauderdale	Florida	33311	South	OFF-ST-1000	Office Supplies	Storage	Eldon Fold 'N	22,368	2	0,2	2,5164	286469,971	0,00%
6	CA-2018-115812	09/06/2018	14/06/2018	Standard	Cl: BH-11710	Brosina Hoff	Consumer	United States	Los Angeles	California	90032	West	FUR-FU-1000	Furniture	Furnishings	Eldon Express	48,86	7	0	14,1694	286469,971	0,00%
7	CA-2018-115812	09/06/2018	14/06/2018	Standard	Cl: BH-11710	Brosina Hoff	Consumer	United States	Los Angeles	California	90032	West	OFF-AR-1000	Office Supplies	Art	Newell 322	7,28	4	0	1,9656	286469,971	0,00%
8	CA-2018-115812	09/06/2018	14/06/2018	Standard	Cl: BH-11710	Brosina Hoff	Consumer	United States	Los Angeles	California	90032	West	TEC-PH-1000	Technology	Phones	Mitel 5320 IP	907,152	6	0,2	90,7152	286469,971	0,03%
9	CA-2018-115812	09/06/2018	14/06/2018	Standard	Cl: BH-11710	Brosina Hoff	Consumer	United States	Los Angeles	California	90032	West	OFF-BI-1000	Office Supplies	Binders	DXL Angle-Vi	18,504	3	0,2	5,7825	286469,971	0,00%
10	CA-2018-115812	09/06/2018	14/06/2018	Standard	Cl: BH-11710	Brosina Hoff	Consumer	United States	Los Angeles	California	90032	West	OFF-AP-1000	Office Supplies	Appliances	Belkin F5C20	114,9	5	0	34,47	286469,971	0,01%
11	CA-2018-115812	09/06/2018	14/06/2018	Standard	Cl: BH-11710	Brosina Hoff	Consumer	United States	Los Angeles	California	90032	West	FUR-TA-1000	Furniture	Tables	Chromcraft R	1706,184	9	0,2	85,3092	286469,971	0,03%
12	CA-2018-115812	09/06/2018	14/06/2018	Standard	Cl: BH-11710	Brosina Hoff	Consumer	United States	Los Angeles	California	90032	West	TEC-PH-1000	Technology	Phones	Konftel 250 C	911,424	4	0,2	68,3568	286469,971	0,02%
13	CA-2021-114412	15/04/2021	20/04/2021	Standard	Cl: AA-10480	Andrew Allen	Consumer	United States	Concord	North Carolina	28027	South	OFF-PA-1000	Office Supplies	Paper	Xerox 1967	15,552	3	0,2	5,4432	286469,971	0,00%
14	CA-2020-161389	05/12/2020	10/12/2020	Standard	Cl: IM-15070	Irene Maddo	Consumer	United States	Seattle	Washington	98103	West	OFF-BI-1000	Office Supplies	Binders	Fellowes PB2	407,976	3	0,2	132,5922	286469,971	0,05%
15	US-2019-118983	22/11/2019	26/11/2019	Standard	Cl: HP-14815	Harold Pawla	Home Office	United States	Fort Worth	Texas	76106	Central	OFF-AP-1000	Office Supplies	Appliances	Holmes Repla	68,81	5	0,8	-123,858	286469,971	-0,04%
16	US-2019-118983	22/11/2019	26/11/2019	Standard	Cl: HP-14815	Harold Pawla	Home Office	United States	Fort Worth	Texas	76106	Central	OFF-BI-1000	Office Supplies	Binders	Storex DuraT	2,544	3	0,8	-3,816	286469,971	0,00%
17	CA-2018-105893	11/11/2018	18/11/2018	Standard	Cl: PK-19075	Pete Kriz	Consumer	United States	Madison	Wisconsin	53711	Central	OFF-ST-1000	Office Supplies	Storage	Stur-D-Stor Sh	665,88	6	0	13,3176	286469,971	0,00%
18	CA-2018-167164	13/05/2018	15/05/2018	Second Class	AG-10270	Alejandro Gru	Consumer	United States	West Jordan	Utah	84084	West	OFF-ST-1000	Office Supplies	Storage	Fellowes Sup	55,5	2	0	9,99	286469,971	0,00%
19	CA-2018-143336	27/08/2018	01/09/2018	Second Class	ZD-21925	Zuschuss Dor	Consumer	United States	San Francisco	California	94109	West	OFF-AR-1000	Office Supplies	Art	Newell 341	8,56	2	0	2,4824	286469,971	0,00%
20	CA-2018-143336	27/08/2018	01/09/2018	Second Class	ZD-21925	Zuschuss Dor	Consumer	United States	San Francisco	California	94109	West	TEC-PH-1000	Technology	Phones	Cisco SPA 50	213,48	3	0,2	16,011	286469,971	0,01%

Column Definitions in the Orders Table:

- Category: The main category under which the product is classified.
- Sub-Category: A more specific classification within the product category.
- Product Name: The name of the product sold.
- Sales: The total revenue generated from each transaction.
- Quantity: The total number of items sold per transaction.
- Discount: The percentage of price reduction applied to the purchased items.
- Profit: The net profit earned from each transaction.
- Fixed LOD: A calculated field using a fixed level of detail expression to ensure consistent aggregation regardless of filters applied in data analysis.
- Profit Contribution: The percentage share of profit contributed by each individual transaction relative to the overall profit.

What's inside the **STATES BAZAAR's** dataset?

Returns

Returned	Order ID
Yes	CA-2021-153822
Yes	CA-2021-129707
Yes	CA-2018-152345
Yes	CA-2019-156440
Yes	US-2021-155999
Yes	CA-2018-157924
Yes	CA-2021-131807
Yes	CA-2020-124527
Yes	CA-2021-135692
Yes	CA-2018-123225
Yes	CA-2021-145772
Yes	US-2018-105137
Yes	CA-2021-101805
Yes	CA-2020-111682
Yes	CA-2021-131492
Yes	CA-2019-104129
Yes	CA-2021-117926
Yes	US-2020-115952
Yes	CA-2019-155761
Yes	CA-2021-100111

Column Definitions in the **Returns** Table:

- Returned: Indicates the status of orders that have been returned or undergone a return process.
- Order ID: A unique code assigned to each transaction.

A total of 296 transactions have been recorded with a Returned status.

What's inside the **STATES BAZAAR's** dataset?

Manager

Manager Name	Region
Anna Andreadi	West
Chuck Magee	East
Kelly Williams	Central
Cassandra Brandow	South

Column Definitions in the Manager Table:

- Manager Name: Indicates the names of managers responsible for overseeing transactions within each respective region.
- Region: The designated geographical area under the responsibility of each manager.

A total of 4 Managers that responsible for each of the Regions.

What's inside the **STATES BAZAAR's** dataset?

Target

Region	Target
Central	600000
East	650000
South	500000
West	700000

Column Definitions in the **TargetTable**:

- Region: Indicates the geographical area associated with each transaction.
- Target: The designated sales target for each region, representing the total revenue goal that must be achieved.

Each region is assigned a different sales target, primarily due to variations in geographical coverage and market potential across regions.



What to do with the dataset?



RFM
Analysis



Sales, Profit, Target
Analysis



Product Performance
Analysis

This dataset provides an opportunity to explore various business metrics. For the purpose of this portfolio, I will focus on three key areas of analysis, each of which will be broken down into reports based on the **Region** variable:

- RFM (Recency, Frequency, Monetary) Analysis
- Sales, Profit, and Target Analysis
- Product Performance



What to do with the dataset?

However, before these analyses can be conducted, the dataset requires several preparatory steps, including:



- **Establishing connections** between multiple sheets using **Power Query**



- **Cleaning** the dataset by addressing missing values



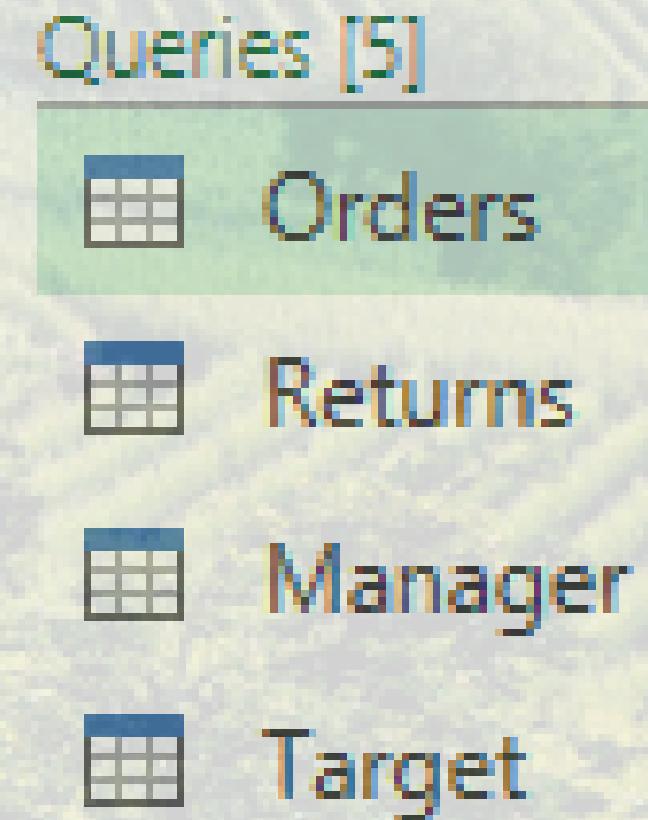
- **Grouping** data by region



- **Creating calculated columns** to enable deeper analytical insights

Finally, an **interactive and dynamic dashboard** will be developed to visualize the data effectively. The dashboard will also incorporate an auto-update mechanism to accommodate newly added data.

Data Preprocessing



In the following pages, I will outline each of the applied steps taken during the data preprocessing phase of this project. The objective is to establish connections across data sources and ensure proper data cleaning, thereby maintaining the overall quality and relevance of the dataset.

The preprocessing will be carried out sequentially, starting with the Orders table, followed by the Returns table, the Manager table, and finally the Regional Target table.

The steps to be undertaken include:

- Data integration from multiple sources
- Removal of duplicate records
- Elimination of missing values
- Data reduction by removing columns irrelevant to the intended analysis
- Outlier detection and removal
- Data type adjustments
- Data transformation
- Creation of conditional columns
- Development of calculated columns
- Implementation of data validation procedures
- Data categorization

Data Preprocessing

Orders Table

APPLIED STEPS	
Source	#
Navigation	#
X Promoted Headers	#
Changed Type	#
Added Custom	#
Reordered Columns	
Inserted Days	#
Reordered Columns1	
Removed Columns	
Renamed Columns	
Removed Columns1	
Merged Queries	#
Expanded Manager	#
Renamed Columns1	
Changed Type1	
Added Conditional Column	#
Merged Queries1	#
Expanded Returns	#

APPLIED STEPS	
X Expanded Returns	#
Added Conditional Column1	#
Removed Columns2	
Added Conditional Column2	#
Reordered Columns2	
Removed Columns3	
Renamed Columns2	
Added Conditional Column3	#
Reordered Columns3	
Removed Columns4	
Added Custom1	#
Reordered Columns4	
Renamed Columns3	
Duplicated Column	
Inserted Year	#
Inserted Quarter	#
Removed Columns5	
Renamed Columns4	

Displayed above are the applied steps performed on the Orders table, each of which will be explained in detail on the following pages.

Data Preprocessing

Orders Table

The screenshot shows the Power BI Data Editor interface. On the left is a preview of the data table with columns: Row ID, Order ID, Order Date, Ship Date, Ship Mode, and Customer ID. The first six rows of data are visible. Above the table is a formula bar containing the M code: `Table.TransformColumnTypes(#"Promoted Headers",{{"Row ID", Int64.Type}, {"Order ID", type text}, {"Order Date", type date}, {"Ship Date", type date}, {"Ship Mode", type text}, {"Customer ID", type text}, {"Customer Name", type text}, {"Segment", type text}, {"Country", type text}, {"City", type text}, {"State", type text}, {"Postal Code", Int64.Type}, {"Region", type text}, {"Product ID", type text}, {"Category", type text}, {"Sub-Category", type text}, {"Product Name", type text}, {"Sales", type number}, {"Quantity", Int64.Type})`. To the right is the 'Query Settings' pane, which includes sections for 'PROPERTIES' (Name: Orders) and 'APPLIED STEPS' (listing 'Source', 'Navigation', 'Promoted Headers', and 'Changed Type', where 'Changed Type' is highlighted in green).

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID
1	CA-2020-152156	08/11/2020	11/11/2020	Second Class	CG-12520
2	CA-2020-152156	08/11/2020	11/11/2020	Second Class	CG-12520
3	CA-2020-138688	12/06/2020	16/06/2020	Second Class	DV-13045
4	US-2019-108966	11/10/2019	18/10/2019	Standard Class	SO-20335
5	US-2019-108966	11/10/2019	18/10/2019	Standard Class	SO-20335
6	CA-2018-115812	09/06/2018	14/06/2018	Standard Class	BH-11710

In the first four steps, I imported the source data from a designated file path on my local device and promoted the first row to headers.

Subsequently, I adjusted the data types of each column to ensure consistency and relevance for the upcoming analysis.

Data Preprocessing

Orders Table

The screenshot shows the Power BI Data Editor interface. On the left, the 'Queries [3]' pane lists 'Orders' (selected), 'Returns', 'Manager', 'Target', and 'Target (2)'. In the center, the 'Orders' query editor shows the following code:

```
# Table.AddColumn(#"Changed Type", "Custom", each [Ship Date]-[Order Date])
```

The preview area displays a table with columns: Row ID, Order ID, Order Date, Ship Date, Ship Mode, and Customer ID. The data consists of 7 rows:

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID
1	CA-2020-152156	08/11/2020	11/11/2020	Second Class	CG-12520
2	CA-2020-152156	08/11/2020	11/11/2020	Second Class	CG-12520
3	CA-2020-138688	12/06/2020	16/06/2020	Second Class	DV-13045
4	US-2019-108966	11/10/2019	18/10/2019	Standard Class	SO-20335
5	US-2019-108966	11/10/2019	18/10/2019	Standard Class	SO-20335
6	CA-2018-115812	09/06/2018	14/06/2018	Standard Class	BH-11710
7	CA-2018-115812	09/06/2018	14/06/2018	Standard Class	BH-11710

On the right, the 'Query Settings' pane shows the 'Name' as 'Orders' and the 'Applied Steps' list, which includes 'Changed Type' and 'Added Custom' (highlighted in green).

In the next step, I created a custom column to calculate the difference between the order received date and the order shipped date.

This was done to capture the recency of each transaction, which will later serve as a key component in the RFM (Recency, Frequency, Monetary) analysis.

Data Preprocessing

Orders Table

The screenshot shows the Power BI Data Editor interface. On the left, the 'Queries [3]' pane lists 'Orders' (selected), 'Returns', 'Manager', 'Target', and 'Target (2)'. In the center, the 'Orders' query editor shows a table with columns: Row ID, Order ID, Order Date, Ship Date, Custom, and Ship Mode. A formula bar at the top contains the DAX code: `# Table.AddColumn(#"Reordered Columns", "Days", each Duration.Days([Custom]), Int64.Type)`. To the right, the 'Query Settings' pane displays the 'PROPERTIES' section with 'Name' set to 'Orders' and the 'APPLIED STEPS' section listing the steps taken: Source, Navigation, Promoted Headers, Changed Type, Added Custom, Reordered Columns, and the final step, 'Inserted Days', which is highlighted with a green border.

Row ID	Order ID	Order Date	Ship Date	Custom	Ship Mode
1	CA-2020-152156	08/11/2020	11/11/2020	3.00:00:00	Second Class
2	CA-2020-152156	08/11/2020	11/11/2020	3.00:00:00	Second Class
3	CA-2020-158688	12/06/2020	16/06/2020	4.00:00:00	Second Class
4	US-2019-108966	11/10/2019	18/10/2019	7.00:00:00	Standard Class
5	US-2019-108966	11/10/2019	18/10/2019	7.00:00:00	Standard Class
6	CA-2018-115812	09/06/2018	14/06/2018	5.00:00:00	Standard Class
7	CA-2018-115812	09/06/2018	14/06/2018	5.00:00:00	Standard Class
8	CA-2018-115812	09/06/2018	14/06/2018	5.00:00:00	Standard Class
9	CA-2018-115812	09/06/2018	14/06/2018	5.00:00:00	Standard Class

The result from the previous step was then converted into a duration in days.

Data Preprocessing

Orders Table

The screenshot shows the Power BI Data Editor interface. On the left, the 'Queries [1]' pane lists 'Orders' as the selected query. The main area displays the 'Orders' table with 13 rows and columns for Order ID, Order Date, Ship Date, Shipping Length, Ship Mode, and Customer ID. Above the table, the M code for the query is shown: `= Table.NestedJoin(#"Removed Columns1", {"Region"}, Manager, {"Region"}, "Manager", JoinKind.LeftOuter)`. To the right, the 'Query Settings' pane shows the query is named 'Orders'. The 'APPLIED STEPS' section lists various transformations applied to the source data.

#	A ⁰ _C Order ID	A ¹ _C Order Date	A ² _C Ship Date	A ³ _C Shipping Length	A ⁴ _C Ship Mode	A ⁵ _C Customer ID
1	CA-2020-152156	08/11/2020	11/11/2020		3 Second Class	CG-12520
2	CA-2020-152156	08/11/2020	11/11/2020		3 Second Class	CG-12520
3	CA-2020-138668	12/06/2020	16/06/2020		4 Second Class	DV-13045
4	US-2019-108966	11/10/2019	18/10/2019		7 Standard Class	SO-20335
5	US-2019-108966	11/10/2019	18/10/2019		7 Standard Class	SO-20335
6	CA-2018-115812	09/06/2018	14/06/2018		5 Standard Class	BH-11710
7	CA-2018-115812	09/06/2018	14/06/2018		5 Standard Class	BH-11710
8	CA-2018-115812	09/06/2018	14/06/2018		5 Standard Class	BH-11710
9	CA-2018-115812	09/06/2018	14/06/2018		5 Standard Class	BH-11710
10	CA-2018-115812	09/06/2018	14/06/2018		5 Standard Class	BH-11710
11	CA-2018-115812	09/06/2018	14/06/2018		5 Standard Class	BH-11710
12	CA-2018-115812	09/06/2018	14/06/2018		5 Standard Class	BH-11710
13	CA-2021-114412	15/04/2021	20/04/2021		5 Standard Class	AA-10480

This step performs a left outer join between the main table and the 'Manager' table using the 'Region' column as the key. It retains all rows from the main table and appends matching rows from 'Manager' as a new column named 'Manager'. If no corresponding match is found in 'Manager', the result will include null values in the appended column for those rows.

Data Preprocessing

Orders Table

The screenshot shows the Power BI Data Editor interface. On the left, the 'Queries [5]' pane lists 'Orders' (selected), 'Returns', 'Manager', 'Target', and 'Target (2)'. The main area displays the 'Orders' table with the following schema:

	A ^R C Order ID	A ^R C Order Date	A ^R C Ship Date	A ^R C Shipping Length	A ^R C Ship Mode	A ^R C Customer ID
1	CA-2020-152156	08/11/2020	11/11/2020		3 Second Class	OG-12520
2	CA-2020-152156	08/11/2020	11/11/2020		3 Second Class	OG-12520
3	US-2019-108966	11/10/2019	18/10/2019		7 Standard Class	SO-20335
4	US-2019-108966	11/10/2019	18/10/2019		7 Standard Class	SO-20335
5	CA-2020-138688	12/06/2020	16/06/2020		4 Second Class	DV-13045
6	CA-2018-115812	09/06/2018	14/06/2018		5 Standard Class	BH-11710
7	CA-2018-115812	09/06/2018	14/06/2018		5 Standard Class	BH-11710
8	CA-2018-115812	09/06/2018	14/06/2018		5 Standard Class	BH-11710
9	CA-2018-115812	09/06/2018	14/06/2018		5 Standard Class	BH-11710
10	CA-2018-115812	09/06/2018	14/06/2018		5 Standard Class	BH-11710
11	CA-2018-115812	09/06/2018	14/06/2018		5 Standard Class	BH-11710
12	CA-2018-115812	09/06/2018	14/06/2018		5 Standard Class	BH-11710
13	CA-2021-114412	15/04/2021	20/04/2021		5 Standard Class	AA-10480
14	CA-2020-161389	05/12/2020	10/12/2020		5 Standard Class	IM-15070

The formula bar at the top shows the DAX query: `= Table.ExpandTableColumn(#"Merged Queries", "Manager", {"Manager Name"}, {"Manager.Manager Name"})`. The 'Query Settings' pane on the right shows the 'Name' is 'Orders' and the 'Applied Steps' include 'Expanded Manager'.

This step expands the previously merged column named "Manager" in the table "Merged Queries". Specifically, it extracts the "Manager Name" field from the nested tables in the "Manager" column and adds it as a new column called "Manager.Manager Name" in the main table.

Basically, it takes the matching manager information (which was nested inside the join) and brings it out as a separate, easy-to-access column.

Data Preprocessing

Orders Table

The screenshot shows the Power BI Data Editor interface. On the left, the 'Queries [5]' pane lists 'Orders' as the selected query. The main area displays the 'Orders' table with 16 rows of data. The columns are: Order ID, Order Date, Ship Date, Shipping Length, Ship Mode, and Customer ID. The 'Applied Steps' pane on the right details the data transformation process, including 'Changed Type' for the 'Sales' and 'Profit' columns.

	Order ID	Order Date	Ship Date	Shipping Length	Ship Mode	Customer ID
1	CA-2020-152156	08/11/2020	11/11/2020	3	Second Class	CG-12520
2	CA-2020-152156	08/11/2020	11/11/2020	3	Second Class	CG-12520
3	US-2019-108966	11/10/2019	18/10/2019	7	Standard Class	SO-20335
4	US-2019-108966	11/10/2019	18/10/2019	7	Standard Class	SO-20335
5	CA-2020-138688	12/06/2020	16/06/2020	4	Second Class	DV-13045
6	CA-2018-115812	09/06/2018	14/06/2018	5	Standard Class	BH-11710
7	CA-2018-115812	09/06/2018	14/06/2018	5	Standard Class	BH-11710
8	CA-2018-115812	09/06/2018	14/06/2018	5	Standard Class	BH-11710
9	CA-2018-115812	09/06/2018	14/06/2018	5	Standard Class	BH-11710
10	CA-2018-115812	09/06/2018	14/06/2018	5	Standard Class	BH-11710
11	CA-2018-115812	09/06/2018	14/06/2018	5	Standard Class	BH-11710
12	CA-2018-115812	09/06/2018	14/06/2018	5	Standard Class	BH-11710
13	CA-2021-114412	15/04/2021	20/04/2021	5	Standard Class	AA-10480
14	CA-2020-161389	05/12/2020	10/12/2020	5	Standard Class	IM-15070
15	US-2019-118983	22/11/2019	26/11/2019	4	Standard Class	HP-14815
16	US-2019-118983	22/11/2019	26/11/2019	4	Standard Class	HP-14815

I converted the "Sales" and "Profit" columns to currency format, and the "Discount" column to percentage format. This ensures that the values in these columns are interpreted and displayed correctly according to their respective data types.

Data Preprocessing

Orders Table

The screenshot shows the Power BI Data Editor interface. On the left, the 'Queries' pane lists 'Orders' as the selected query. The main area displays the 'Orders' table with columns: Order ID, Order Date, Ship Date, Shipping Length, Ship Mode, and Customer ID. A new column, 'Profit Status', has been added, which categorizes each row as 'Profit' or 'Loss' based on the value in the Profit column. The 'Applied Steps' pane on the right details the steps taken, including 'Changed Type1' for the 'Profit Status' column.

	Order ID	Order Date	Ship Date	Shipping Length	Ship Mode	Customer ID
1	CA-2020-152156	08/11/2020	11/11/2020	3	Second Class	CG-12520
2	CA-2020-152156	08/11/2020	11/11/2020	3	Second Class	CG-12520
3	US-2019-108966	11/10/2019	18/10/2019	7	Standard Class	SO-20335
4	US-2019-108966	11/10/2019	18/10/2019	7	Standard Class	SO-20335
5	CA-2020-138688	12/06/2020	16/06/2020	4	Second Class	DV-13045
6	CA-2018-115812	09/06/2018	14/06/2018	5	Standard Class	BH-11710
7	CA-2018-115812	09/06/2018	14/06/2018	5	Standard Class	BH-11710
8	CA-2018-115812	09/06/2018	14/06/2018	5	Standard Class	BH-11710
9	CA-2018-115812	09/06/2018	14/06/2018	5	Standard Class	BH-11710
10	CA-2018-115812	09/06/2018	14/06/2018	5	Standard Class	BH-11710
11	CA-2018-115812	09/06/2018	14/06/2018	5	Standard Class	BH-11710
12	CA-2018-115812	09/06/2018	14/06/2018	5	Standard Class	BH-11710
13	CA-2021-114412	15/04/2021	20/04/2021	5	Standard Class	AA-10480
14	CA-2020-161389	08/13/2020	10/13/2020	5	Standard Class	IM-15070
15	US-2019-118983	22/11/2019	26/11/2019	4	Standard Class	HP-14815
16	US-2019-118983	22/11/2019	26/11/2019	4	Standard Class	HP-14815
17	CA-2018-105893	11/11/2018	16/11/2018	7	Standard Class	PK-19075

This step adds a new column called "Profit Status". For each row, it checks the value in the "Profit" column: if the profit is greater than zero, it assigns the label "Profit"; otherwise, it assigns "Loss". This categorizes each row based on whether it represents a profit or a loss.

Data Preprocessing

Orders Table

The screenshot shows the Power BI Data Editor interface. On the left, the 'Queries' pane lists 'Orders' (selected), 'Returns', 'Manager', 'Target', and 'Target (2)'. The main area displays a table with columns: Order ID, Order Date, Ship Date, Shipping Length, Ship Mode, and Customer ID. A tooltip indicates the query is a 'Table.NestedJoin' operation. The table data is as follows:

	Order ID	Order Date	Ship Date	Shipping Length	Ship Mode	Customer ID
1	CA-2020-152156	08/11/2020	11/11/2020		3 Second Class	OG-12520
2	CA-2020-152156	08/11/2020	11/11/2020		3 Second Class	OG-12520
3	US-2019-108966	11/10/2019	18/10/2019		7 Standard Class	SO-20335
4	US-2019-108966	11/10/2019	18/10/2019		7 Standard Class	SO-20335
5	CA-2020-138688	12/06/2020	16/06/2020		4 Second Class	OV-13045
6	CA-2018-115812	09/06/2018	14/06/2018		5 Standard Class	BH-11710
7	CA-2018-115812	09/06/2018	14/06/2018		5 Standard Class	BH-11710
8	CA-2018-115812	09/06/2018	14/06/2018		5 Standard Class	BH-11710
9	CA-2018-115812	09/06/2018	14/06/2018		5 Standard Class	BH-11710
10	CA-2018-115812	09/06/2018	14/06/2018		5 Standard Class	BH-11710
11	CA-2018-115812	09/06/2018	14/06/2018		5 Standard Class	BH-11710
12	CA-2018-115812	09/06/2018	14/06/2018		5 Standard Class	BH-11710
13	CA-2021-114412	15/04/2021	20/04/2021		5 Standard Class	AA-10480
14	CA-2020-161389	05/12/2020	10/12/2020		5 Standard Class	IM-15070
15	CA-2020-111682	17/06/2020	18/06/2020		1 First Class	TB-21055
16	CA-2020-111682	17/06/2020	18/06/2020		1 First Class	TB-21055
17	CA-2020-111682	17/06/2020	18/06/2020		1 First Class	TB-21055
18	CA-2020-111682	17/06/2020	18/06/2020		1 First Class	TB-21055

The 'Query Settings' pane on the right shows the query name is 'Orders' and lists applied steps including 'Merged Queries1'.

This step performs a full outer join between the main table and the "Returns" table using the "Order ID" column as the matching key in both tables. The result includes all rows from both tables—matching records are combined, while non-matching records from either table are preserved, with null values in the fields where no match exists. The joined data from the "Returns" table is stored in a new column named "Returns".

Data Preprocessing

Orders Table

The screenshot shows the Power BI Data Editor interface. On the left, there's a sidebar with 'Queries [1]' containing 'Orders' (selected), 'Returns', 'Manager', 'Target', and 'Target (2)'. The main area displays a table with the following data:

	Order ID	Order Date	Ship Date	Shipping Length	Ship Mode	Customer ID
1	CA-2020-152156	08/11/2020	11/11/2020	123	Second Class	CG-12520
2	CA-2020-152156	08/11/2020	11/11/2020	123	Second Class	CG-12520
3	US-2019-158946	11/10/2019	18/10/2019	123	Standard Class	SO-20335

A tooltip above the table reads: '# Table.ExpandTableColumn(#"Merged Queries1", "Returns", {"Returned"}, {"Returns.Returned"})'. To the right, the 'Query Settings' pane shows 'Name: Orders' and 'Applied Steps: Expanded Returns'.

This step expands the "Returns" column, which contains nested tables resulting from a previous join, in the main table. It extracts the "Returned" field from each nested table and adds it as a new column named "Returns.Returned". This makes the return status data directly accessible in the main table.

Data Preprocessing

Orders Table

The screenshot shows the Power BI Data Editor interface. On the left, the 'Queries' pane lists 'Orders' as the active query. In the center, a preview of the 'Orders' table is shown with columns: Order ID, Order Date, Ship Date, Shipping Length, Ship Mode, and Customer ID. A formula bar at the top contains DAX code: `Table.AddColumn("Expanded Returns", "Return Status", each if ([Returns.Returned]) = "Yes" then "Returned" else if ([Returns.Returned]) = null then "Not Returned" else null)`. To the right, the 'Query Settings' pane shows the query name is 'Orders'. The 'APPLIED STEPS' section lists two steps: 'Expanded Returns' and 'Added Conditional Column1'.

Order ID	Order Date	Ship Date	Shipping Length	Ship Mode	Customer ID
CA-2020-152156	08/11/2020	11/11/2020		Second Class	CG-12530
CA-2020-152156	08/11/2020	11/11/2020		Second Class	CG-12530
US-2019-108946	11/10/2019	18/10/2019		Standard Class	SO-20335
US-2019-108946	11/10/2019	18/10/2019		Standard Class	SO-20335

This step adds a new column named "Return Status" to the main table. For each row, it checks the value in the "Returns.Returned" column:

- If the value is "Yes", it assigns "Returned".
- If the value is null, it assigns "Not Returned".
- Otherwise, it assigns null.

This logic is used to categorize orders based on their return status, making it easier to interpret whether an order was returned or not.

Data Preprocessing

Orders Table

The screenshot shows the Power BI Data Editor interface. On the left, the 'Queries [1]' pane lists 'Orders' as the selected query. In the center, the 'Table.AddColumn("Removed Columns2", "SL trial", each if [Shipping Length] > 100 then 0 else [Shipping Length])' step is displayed above a preview of the 'Orders' table. The table has columns: Order ID, Order Date, Ship Date, Shipping Length, Ship Mode, and Customer ID. The preview shows 6 rows of data. On the right, the 'Query Settings' pane shows the query name is 'Orders'. The 'APPLIED STEPS' section lists: Expanded Returns, Added Conditional Column1, Removed Columns2, and the current step, Added Conditional Column2.

Order ID	Order Date	Ship Date	Shipping Length	Ship Mode	Customer ID
CA-2020-152156	08/11/2020	11/11/2020	3	Second Class	CG-12520
CA-2020-152156	08/11/2020	11/11/2020	3	Second Class	CG-12520
US-2019-108966	11/10/2019	18/10/2019	7	Standard Class	SO-20335
US-2019-108966	11/10/2019	18/10/2019	7	Standard Class	SO-20335
CA-2020-134688	12/06/2020	16/06/2020	4	Second Class	DV-13045
CA-2018-115812	09/06/2018	14/06/2018	5	Standard Class	BH-11710

This step adds a new column named "SL trial" to the main table. For each row, it evaluates the value in the "Shipping Length" column:

- If the shipping length is greater than 100, it assigns a value of 0.
- Otherwise, it retains the original Shipping Length value.

This transformation is used to cap or filter unusually high shipping length values for further analysis.

Data Preprocessing

Orders Table

The screenshot shows the Power BI Data Editor interface. On the left, the 'Queries' pane lists 'Orders' as the selected query. In the center, a code editor window displays the M code for creating a new column:

```
= Table.AddColumn(#"Renamed Columns2", "Shipping Time", each if [Shipping Length] < 0 then 0 else [Shipping Length])
```

Below the code is a table preview showing 10 rows of data. The columns are: Order ID, Order Date, Ship Date, Shipping Length, Ship Mode, and Customer ID. The data includes various order IDs, dates ranging from 2018 to 2020, and shipping modes like Second Class and Standard Class.

On the right, the 'Query Settings' pane shows the query name is 'Orders'. The 'APPLIED STEPS' section is expanded, listing the following steps:

- Expanded Returns
- Added Conditional Column1
- Removed Columns2
- Added Conditional Column2
- Reordered Columns2
- Removed Columns3
- Renamed Columns2
- Added Conditional Column3

This step adds a new column named "Shipping Time" to the main table. For each row, it checks the value in the "Shipping Length" column:

- If the value is less than 0, it replaces it with 0.
- Otherwise, it retains the original Shipping Length value.

This logic is used to correct invalid or negative shipping durations by resetting them to zero, ensuring data consistency for further analysis.

Data Preprocessing

Orders Table

The screenshot shows the Power BI Data Editor interface. On the left, the 'Queries [1]' pane lists 'Orders' as the selected query. The main area displays the 'Orders' table with the following columns: Order ID, Order Date, Ship Date, Shipping Time, Ship Mode, and Customer ID. A new column, 'Shipping Time(hours)', has been added to the table. The 'Query Settings' pane on the right shows the query name is 'Orders' and lists the applied steps, including 'Added Custom1'.

	Order ID	Order Date	Ship Date	Shipping Time	Ship Mode	Customer ID
1	CA-2020-152156	08/11/2020	11/11/2020	3	Second Class	CG-12520
2	CA-2020-152156	08/11/2020	11/11/2020	3	Second Class	CG-12520
3	US-2019-108966	11/10/2019	18/10/2019	7	Standard Class	SO-20335
4	US-2019-108966	11/10/2019	18/10/2019	7	Standard Class	SO-20335
5	CA-2020-138688	12/06/2020	16/06/2020	4	Second Class	DV-13045
6	CA-2018-115812	09/06/2018	14/06/2018	5	Standard Class	BH-11710
7	CA-2018-115812	09/06/2018	14/06/2018	5	Standard Class	BH-11710
8	CA-2018-115812	09/06/2018	14/06/2018	5	Standard Class	BH-11710
9	CA-2018-115812	09/06/2018	14/06/2018	5	Standard Class	BH-11710
10	CA-2018-115812	09/06/2018	14/06/2018	5	Standard Class	BH-11710
11	CA-2018-115812	09/06/2018	14/06/2018	5	Standard Class	BH-11710
12	CA-2018-115812	09/06/2018	14/06/2018	5	Standard Class	BH-11710
13	CA-2021-114412	15/04/2021	20/04/2021	5	Standard Class	AA-10480

This step adds a new column named "Shipping Time(hours)" to the main table. For each row, it multiplies the value in the "Shipping Time" column by 24, converting the shipping time from days to hours. This provides a more granular representation of the shipping duration.

Data Preprocessing

Orders Table

The screenshot shows a data processing interface with the following components:

- Overviews [3]**: A sidebar with a tree view showing three main categories: Orders, Returns, Manager, Target, and Target (2). The Orders node is currently selected.
- Query Settings**: A panel on the right containing:
 - PROPERTIES**: A section for naming the query, with "Name" set to "Orders".
 - APPLIED STEPS**: A list of 20 steps taken during the preprocessing, including:
 - Expanded Returns
 - Added Conditional Column1
 - Removed Columns2
 - Added Conditional Column2
 - Reordered Columns2
 - Removed Columns3
 - Renamed Columns2
 - Added Conditional Column3
 - Reordered Columns3
 - Removed Columns4
 - Added Custom1
 - Reordered Columns4
 - Renamed Columns3
 - Duplicated Column
 - Inserted Year
- Table Editor**: The main area displaying the Orders table with the following columns and data:

	Order ID	Order Date	Ship Date	Shipping Time(days)	Shipping Time(hours)	Ship Mode
1	CA-2020-152156	08/11/2020	11/11/2020	3	72	Second Class
2	CA-2020-152156	08/11/2020	11/11/2020	3	72	Second Class
3	US-2019-108966	11/10/2019	18/10/2019	7	168	Standard
4	US-2019-108966	11/10/2019	18/10/2019	7	168	Standard
5	CA-2020-138688	12/06/2020	16/06/2020	4	96	Second Class
6	CA-2018-115812	09/06/2018	14/06/2018	5	120	Standard
7	CA-2018-115812	09/06/2018	14/06/2018	5	120	Standard
8	CA-2018-115812	09/06/2018	14/06/2018	5	120	Standard
9	CA-2018-115812	09/06/2018	14/06/2018	5	120	Standard
10	CA-2018-115812	09/06/2018	14/06/2018	5	120	Standard
11	CA-2018-115812	09/06/2018	14/06/2018	5	120	Standard
12	CA-2018-115812	09/06/2018	14/06/2018	5	120	Standard
13	CA-2021-114412	15/04/2021	20/04/2021	5	120	Standard
14	CA-2020-161389	05/12/2020	10/12/2020	5	120	Standard
15	CA-2020-111682	17/06/2020	18/06/2020	1	24	First Class
16	CA-2020-111682	17/06/2020	18/06/2020	1	24	First Class
17	CA-2020-111682	17/06/2020	18/06/2020	1	24	First Class

This step adds a new column named "Year" to the main table. For each row, it extracts the year component from the "Order Date - Copy" column using the Date.Year function. The extracted year is stored as a 64-bit integer (Int64.Type), ensuring it is recognized as a numeric data type suitable for analysis and sorting.

Data Preprocessing

Orders Table

The screenshot shows the Power BI Data Editor interface. On the left, the 'Queries' pane lists 'Orders' as the selected query. The main area displays the 'Orders' table with columns: Order ID, Order Date, Ship Date, Shipping Time(days), Shipping Time(hours), and Ship Mode. A new column, 'Inserted Quarter', is added to the right of the table. The formula bar at the top shows the DAX code: `= Table.AddColumn(#"Inserted Year", "Quarter", each Date.QuarterOfYear([#"Order Date - Copy"]), Int64.Type)`. The 'Query Settings' pane on the right shows the query name is 'Orders'. The 'APPLIED STEPS' pane lists the steps taken to transform the data, including 'Inserted Year' and 'Inserted Quarter'.

Order ID	Order Date	Ship Date	Shipping Time(days)	Shipping Time(hours)	Ship Mode	Inserted Quarter
CA-2020-152156	08/11/2020	11/11/2020	3	72	Second Class	3
CA-2020-152156	08/11/2020	11/11/2020	3	72	Second Class	3
US-2019-108946	11/10/2019	18/10/2019	7	168	Standard	3
US-2019-108946	11/10/2019	18/10/2019	7	168	Standard	3
CA-2020-138648	12/06/2020	16/06/2020	4	96	Second Class	3
CA-2018-115812	09/06/2018	14/06/2018	5	120	Standard	3
CA-2018-115812	09/06/2018	14/06/2018	5	120	Standard	3
CA-2018-115812	09/06/2018	14/06/2018	5	120	Standard	3
CA-2018-115812	09/06/2018	14/06/2018	5	120	Standard	3
CA-2018-115812	09/06/2018	14/06/2018	5	120	Standard	3
CA-2018-115812	09/06/2018	14/06/2018	5	120	Standard	3
CA-2018-115812	09/06/2018	14/06/2018	5	120	Standard	3
CA-2021-114412	15/04/2021	20/04/2021	5	120	Standard	2
CA-2020-161389	05/12/2020	10/12/2020	5	120	Standard	4
CA-2020-111682	17/06/2020	18/06/2020	1	24	First Class	3
CA-2020-111682	17/06/2020	18/06/2020	1	24	First Class	3
CA-2020-111682	17/06/2020	18/06/2020	1	24	First Class	3

This step adds a new column named "Quarter" to the main table. For each row, it calculates the quarter of the year (ranging from 1 to 4) based on the value in the "Order Date - Copy" column using the Date.QuarterOfYear function. The result is stored as a 64-bit integer (Int64.Type), which enables numerical operations and chronological grouping in subsequent analysis.

Data Preprocessing

Returns and Manager Table

The screenshot shows the Power BI Data Editor interface. On the left, the 'Queries [3]' pane lists 'Orders', 'Returns' (selected), 'Manager', and 'Target'. The main area displays the 'Returns' table with columns 'Returned' and 'Order ID'. The data shows 10 rows where 'Returned' is 'Yes' and 'Order ID' is a unique identifier. To the right, the 'Query Settings' pane shows the 'Name' is 'Returns' and the 'APPLIED STEPS' section includes 'Removed Top Rows'.

Returned	Order ID
Yes	CA-2021-155822
Yes	CA-2021-129707
Yes	CA-2018-152345
Yes	CA-2019-156440
Yes	US-2021-155999
Yes	CA-2018-157924
Yes	CA-2021-131807
Yes	CA-2020-124527
Yes	CA-2021-135692
Yes	CA-2018-123125

The screenshot shows the Power BI Data Editor interface. On the left, the 'Queries [3]' pane lists 'Orders', 'Returns' (selected), 'Manager', and 'Target'. The main area displays the 'Manager' table with columns 'Manager Name' and 'Region'. The data shows 4 rows with names like Anna Andreadi, Chuck Magee, Kelly Williams, and Cassandra Brandow, assigned to West, East, Central, and South regions respectively. To the right, the 'Query Settings' pane shows the 'Name' is 'Manager' and the 'APPLIED STEPS' section includes 'Removed Top Rows'.

Manager Name	Region
Anna Andreadi	West
Chuck Magee	East
Kelly Williams	Central
Cassandra Brandow	South

The same data cleaning procedures were applied to two additional tables: the Returns table and the Managers table.

This step involved the removal of duplicates, outliers, and empty rows. Data validation was also performed to ensure optimal and relevant data processing.

Data Preprocessing

Target Table

The screenshot shows the Power BI Data Editor interface. On the left, the 'Queries [3]' pane lists 'Orders', 'Returns', 'Manager', 'Target' (selected), and 'Target (2)'. In the center, the 'Target' table is displayed with columns 'Region' and 'Target'. The data rows are: 1 Central 600000, 2 East 650000, 3 South 500000, 4 West 700000. Above the table, the M code step 'Table.TransformColumnTypes("Promoted Headers", {"Region", type text}, {"Target", Int64.Type}))' is shown. On the right, the 'Query Settings' pane shows the 'Name' is 'Target' and the 'Applied Steps' list includes 'Source', 'Navigation', 'Promoted Headers', and 'Changed Type' (highlighted in green).

	Region	Target
1	Central	600000
2	East	650000
3	South	500000
4	West	700000

This step changes the data types of specific columns in the Target table. It converts the "Region" column to text data type and the "Target" column to a 64-bit integer (Int64.Type). This ensures that each column has the appropriate data type for accurate analysis and data processing.

Final Structured Dataset

	A	B	C	D	E	F	G	H	I	J	K
1	Order ID	Order Date	Ship Date	Order Recency Days	Shipping Time(days)	Shipping Time(hours)	Ship Mode	Customer ID	Customer Name	Segment	Country
2	CA-2018-103800	03/01/2018	07/01/2018	0	4	96	Standard Class	DP-13000	Darren Powers	Consumer	United States
3	CA-2018-112326	04/01/2018	08/01/2018	0	4	96	Standard Class	PO-19195	Phillina Ober	Home Office	United States
4	CA-2018-112326	04/01/2018	08/01/2018	0	4	96	Standard Class	PO-19195	Phillina Ober	Home Office	United States
5	CA-2018-112326	04/01/2018	08/01/2018	1	4	96	Standard Class	PO-19195	Phillina Ober	Home Office	United States

	L	M	N	O	P	Q	R	S
1	City	State	Postal Code	Region	Product ID	Category	Sub-Category	Product Name
2	Houston	Texas	77095	Central	OFF-PA-10000174	Office Supplies	Paper	Message Book, Wirebound, Four 5 1/2" X 4" Forms/Pg., 200 Dupl. Sets/Book
3	Naperville	Illinois	60540	Central	OFF-LA-10003223	Office Supplies	Labels	Avery 508
4	Naperville	Illinois	60540	Central	OFF-ST-10002743	Office Supplies	Storage	SAFCO Boltless Steel Shelving
5	Naperville	Illinois	60540	Central	OFF-BI-10004094	Office Supplies	Binders	GBC Standard Plastic Binding Systems Combs

	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE
1	Sales	Sales Av	Quantity	Discount	Profit	Fixed LOD Profit	Profit Contribution	Responsible Manager	Profit Status	Return Status	Transaction Year	Transaction Quarter
2	16,448	229,86	2	0,2	5,5512	286469,9713	1,93779E-05	Kelly Williams	Profit	Not Returned	2018	1
3	11,784	229,86	3	0,2	4,2717	286469,9713	1,49115E-05	Kelly Williams	Profit	Not Returned	2018	1
4	272,736	229,86	3	0,2	-64,7748	286469,9713	-0,000226114	Kelly Williams	Loss	Not Returned	2018	1
5	3,54	229,86	2	0,8	-5,487	286469,9713	-1,91538E-05	Kelly Williams	Loss	Not Returned	2018	1

Brief Overview of the Analyses to Be Conducted

All analyses presented in this report are conducted with a **regional** focus, aiming to generate insights specific to each geographical area.



RFM Analysis

This analysis examines the recency, frequency, and monetary value of transactions across regions. Although RFM is commonly used for customer segmentation, in this context it serves as a **diagnostic tool** to understand transactional behavior and patterns at the regional level.

Sales, Profit, and Target Analysis

This section explores the **relationship between sales and profit**, as well as the **comparison between actual sales and predefined sales targets**. The analysis aims to evaluate regional performance in achieving profitability and sales goals.

Product Performance Analysis

This analysis assesses **how different products perform across regions**, considering multiple dimensions such as product popularity (most sought-after items), sales value, profitability, return rates, and shipment activity. The goal is to **identify which products drive success or require strategic attention in each region**.

Backend

Vs. Frontend

After obtaining a well-structured dataset that is ready for in-depth analysis, the reporting will be divided into two main sections.

The first section is the **backend**, which contains pivot tables for each analysis. These tables are connected to slicers for dynamic metric filtering.

The second section is the **frontend**, which consists of an interactive and dynamic dashboard. This dashboard is designed to assist users in interpreting the data and supporting business discussions through relevant visualizations.



Backend

Pivot Tables, Slicers,
Calculations

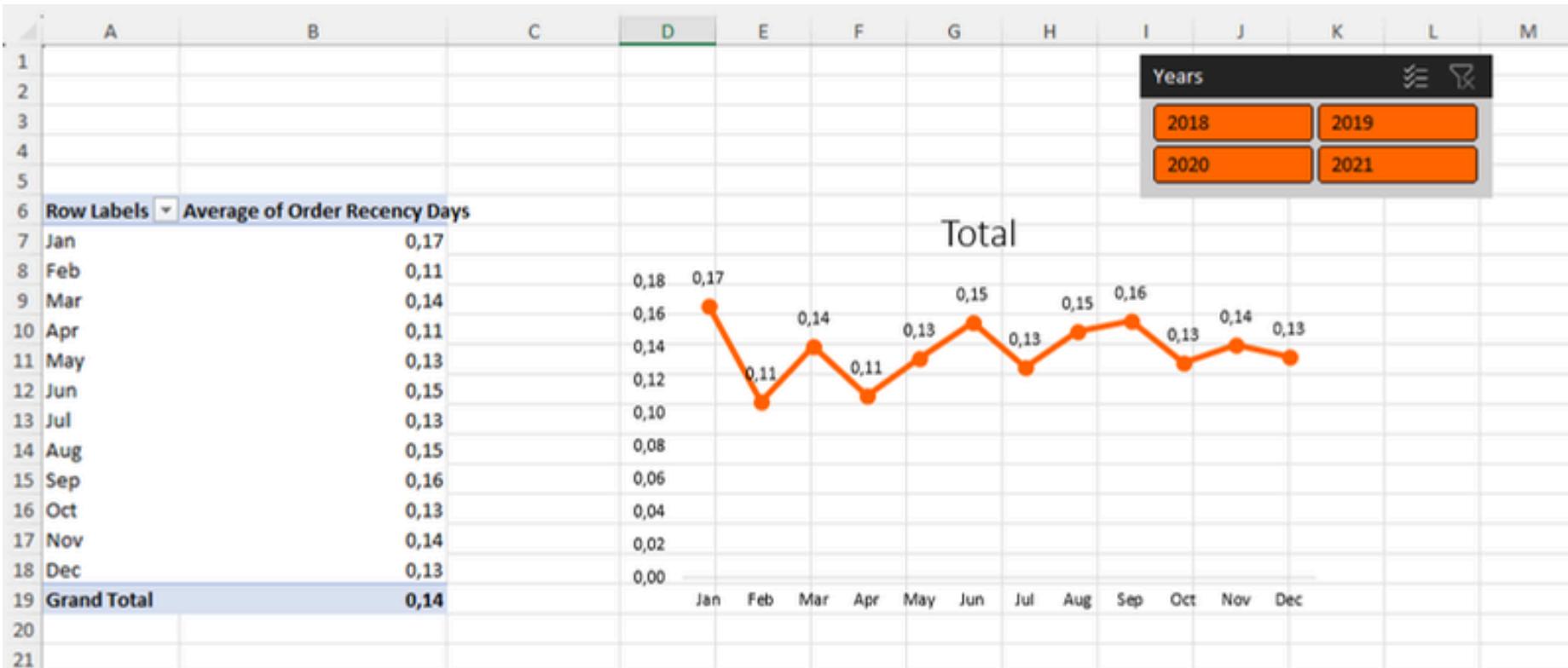


Frontend

Interactive Dashboard,
Charts, Visualizations



The Dashboard



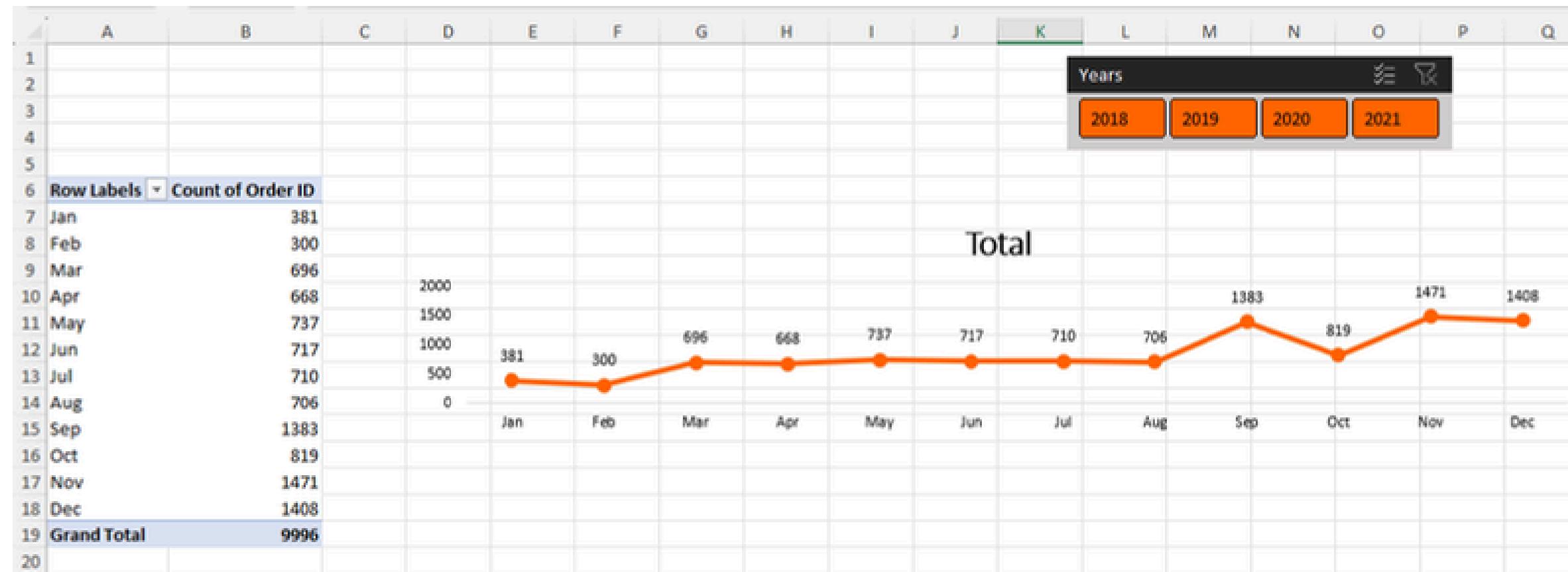
In the backend of the RFM analysis, I constructed a pivot table with months as the X-axis and the average transaction interval (recency) as the measured value.

This represents the average number of days between the most recent transaction and the previous one.

The pivot table was then visualized using a line chart to provide a clearer view of the overall monthly trends.

Next, I applied a regional filter to enable comparisons across different regions, allowing for a more granular understanding of regional transaction patterns.

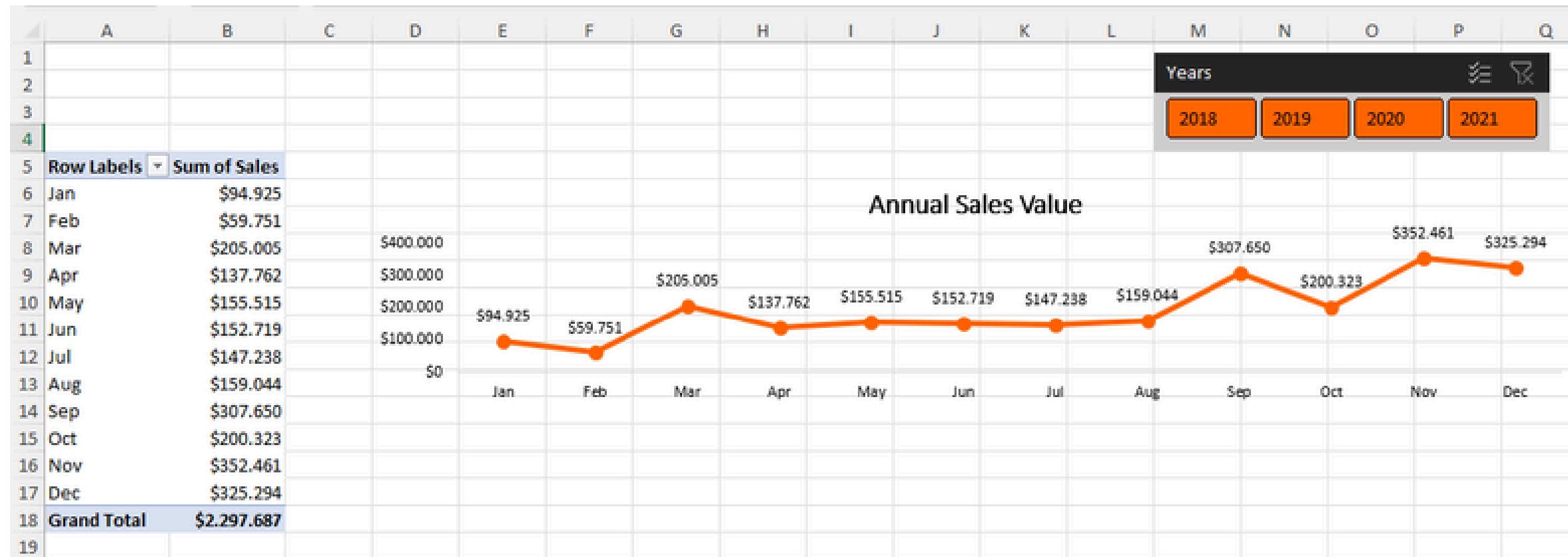
A slicer was also implemented to filter the data by year, making it easier to observe and compare temporal changes within and across regions.



Following the same approach as in the previous step, I created another pivot table—this time using the average daily transaction intensity as the measured value.

The goal of this step is to observe the average number of transactions occurring per day within each month, across different regions.

As before, a regional filter was applied to enable cross-regional comparison, and a year-based slicer was included to facilitate temporal analysis.



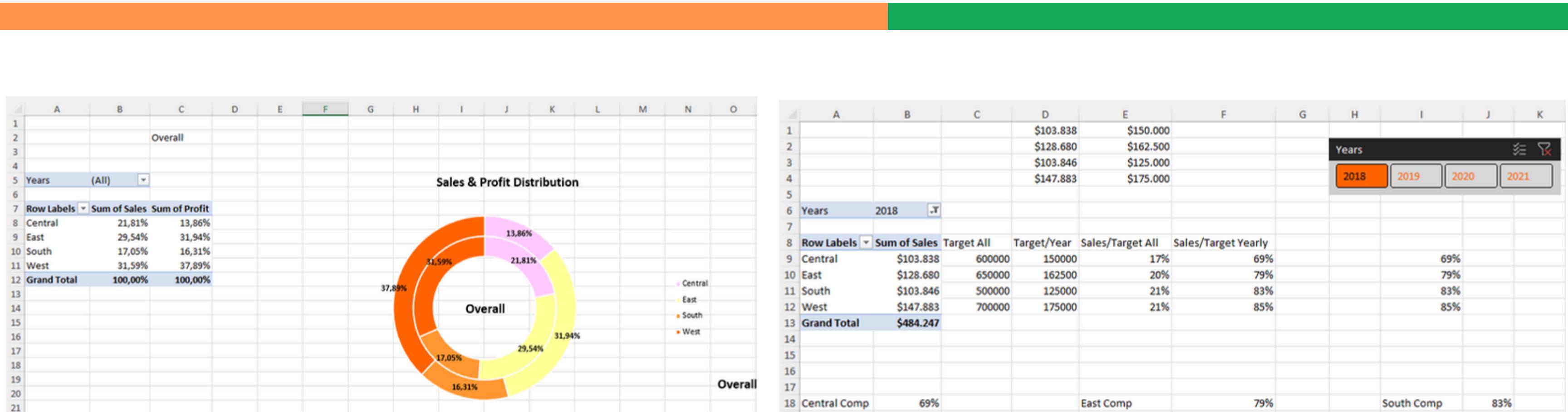
For the monetary component of the analysis, the measured value used is total sales.

The visualization derived from this pivot table provides a comprehensive view of total monthly sales across all regions.

A year-based slicer is applied to allow users to access and compare sales performance on an annual basis.

Backend

Sales, Profit, Target Analysis



The next section focuses on Sales, Profit, and Target analysis.

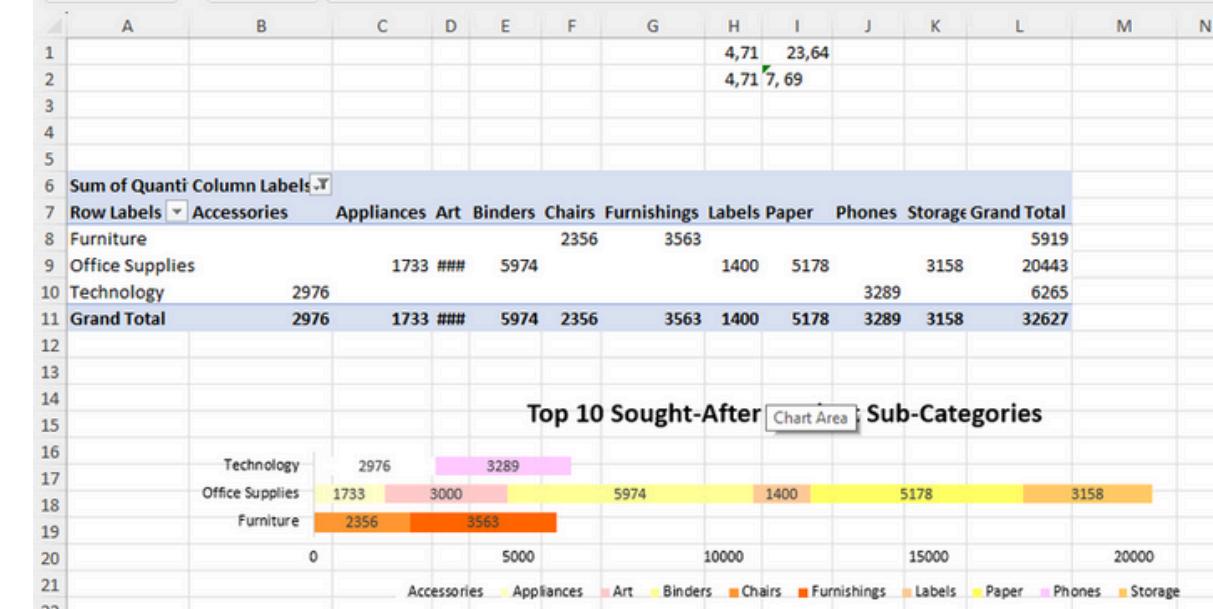
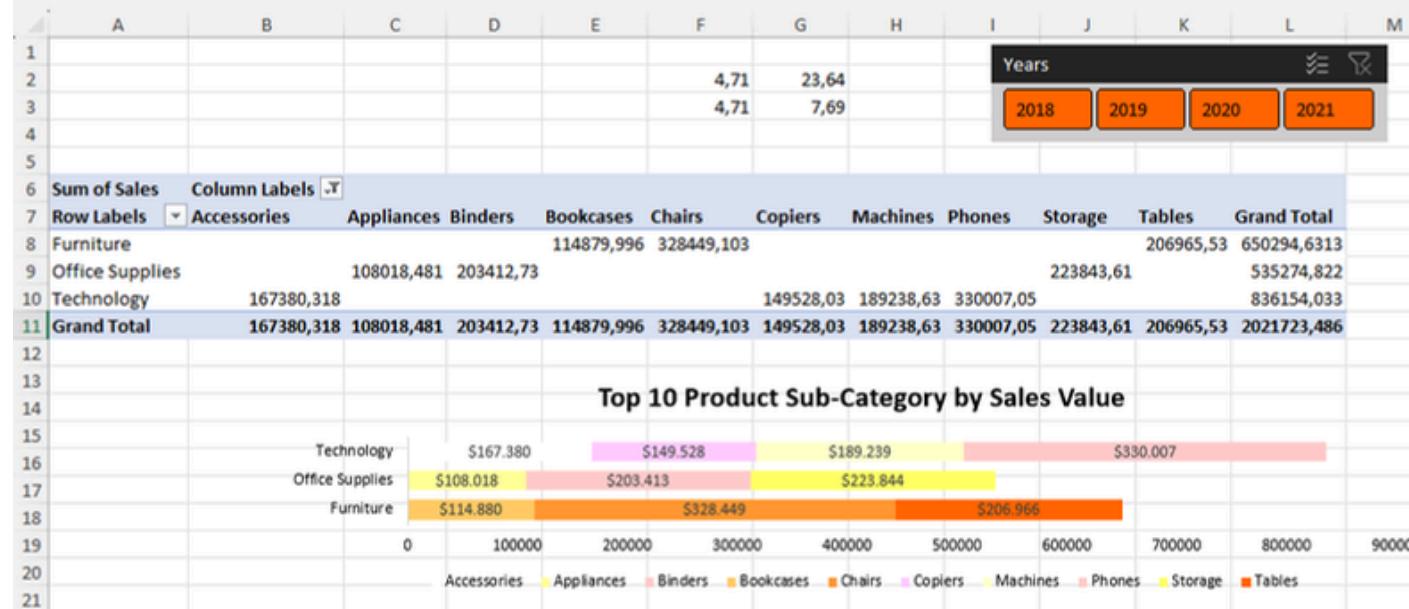
In this part, I created a pivot table to support data visualization, enabling a clear view of the distribution of sales, profit, and target values across different regions.

This analysis is intended to provide a comprehensive understanding of business activity at the regional level, serving as a basis for data-driven decision-making.

A year-based slicer was also implemented to allow users to explore annual changes in sales, profit, and their alignment with pre-established targets.

Backend

Product Performance Analysis

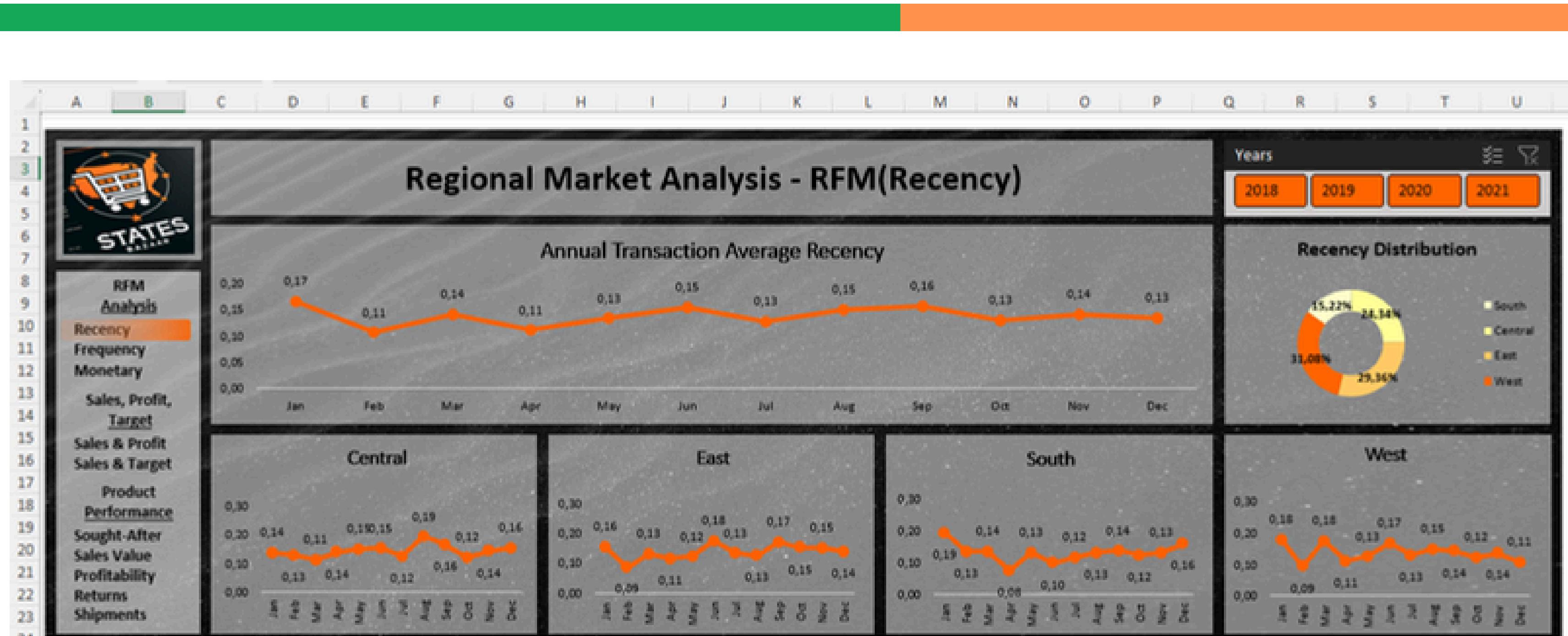


The final section focuses on Product Performance. This analysis aims to provide a comprehensive understanding and clear visibility into the products being sold, as well as the preferred shipping methods associated with each product category.

The discussion covers several key metrics, including:

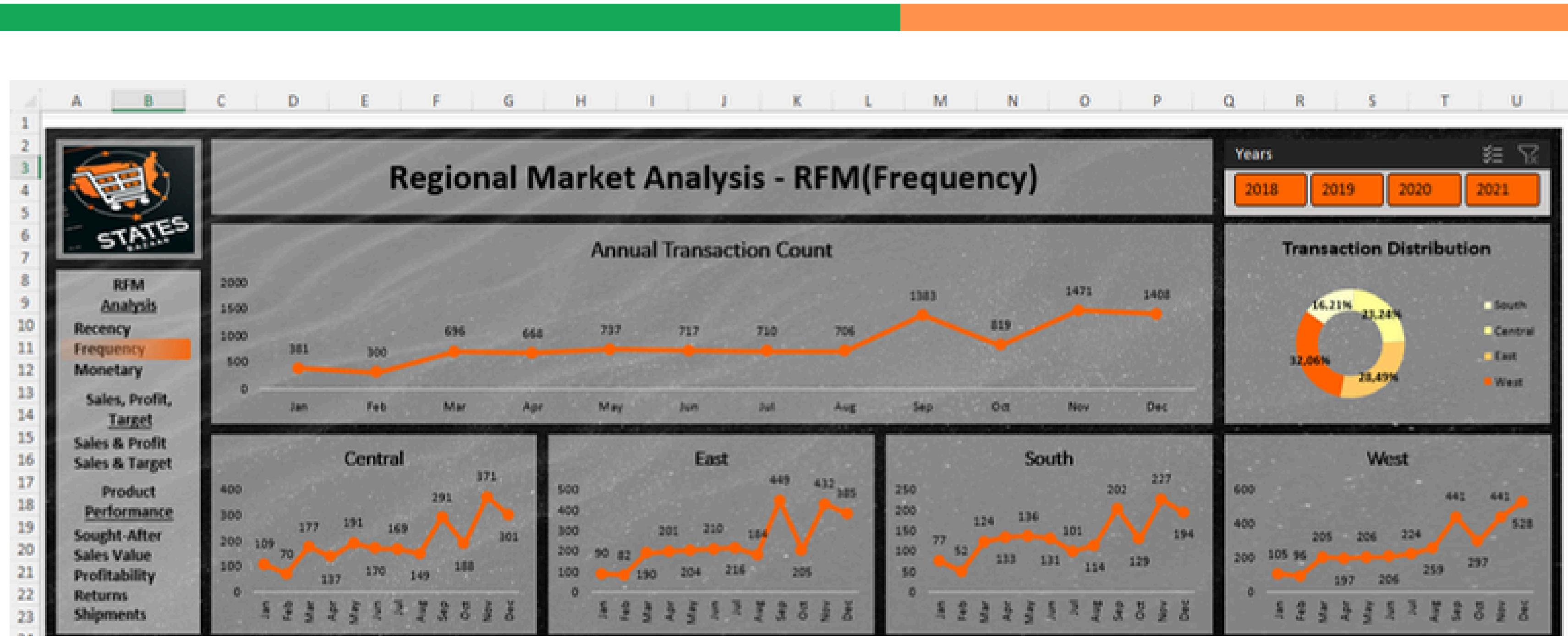
- Best-selling products
- Products with the highest sales value
- Most profitable products
- Most frequently returned products
- Shipping method preferences by product category

By examining these dimensions, the analysis supports deeper insight into product-level dynamics across regions and helps inform more strategic decisions related to inventory, sales, and logistics.



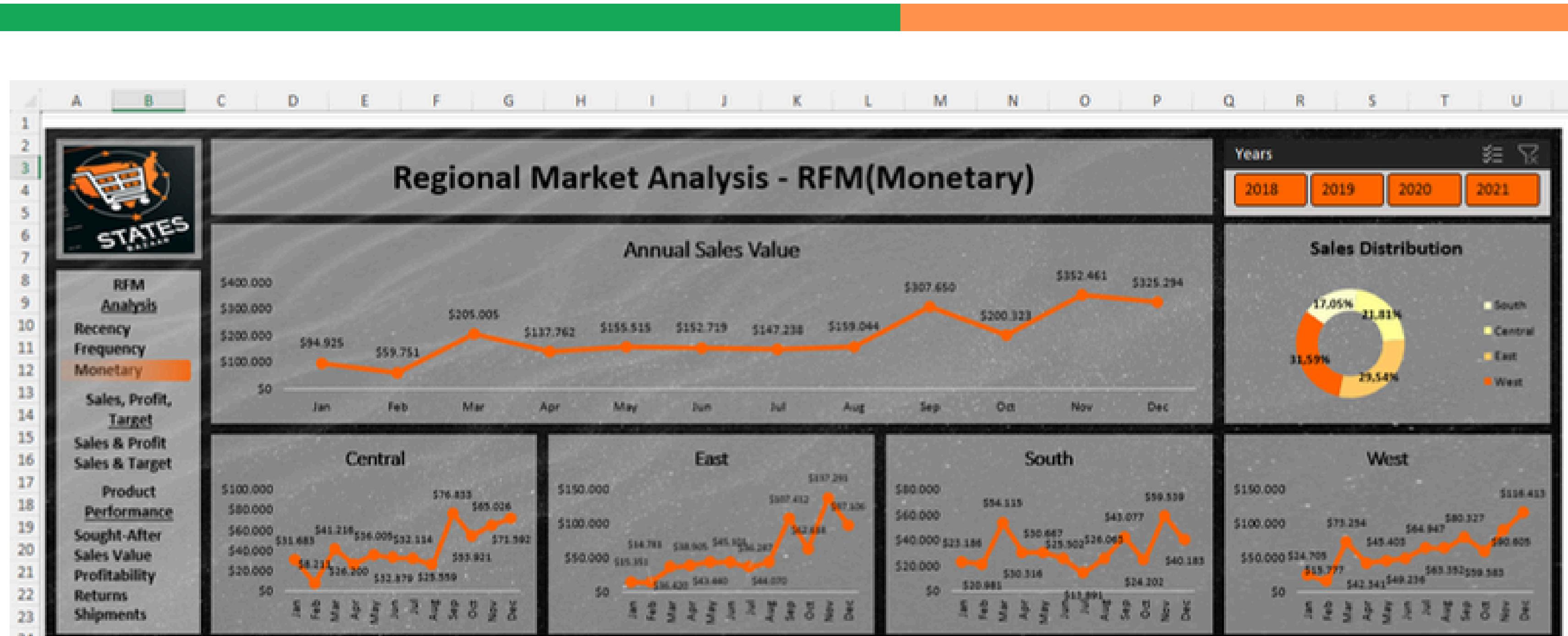
Frontend

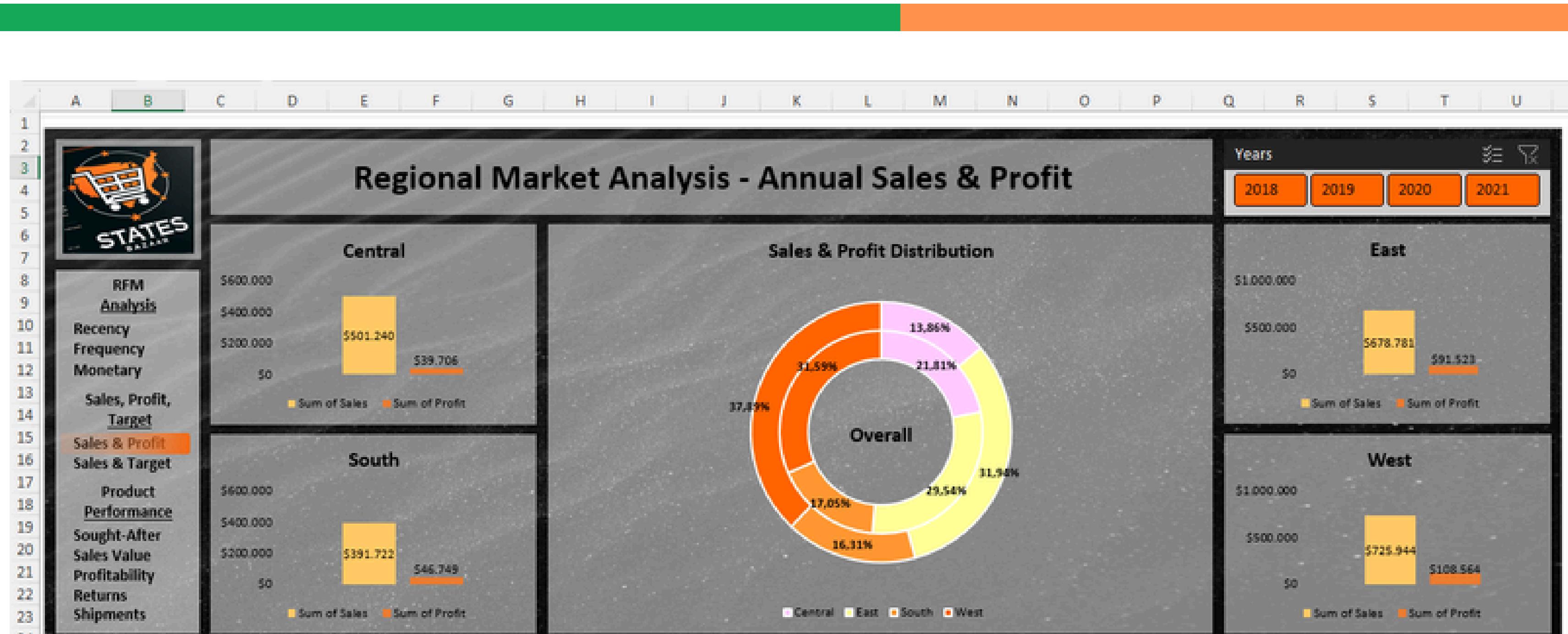
RFM Analysis – Frequency



Frontend

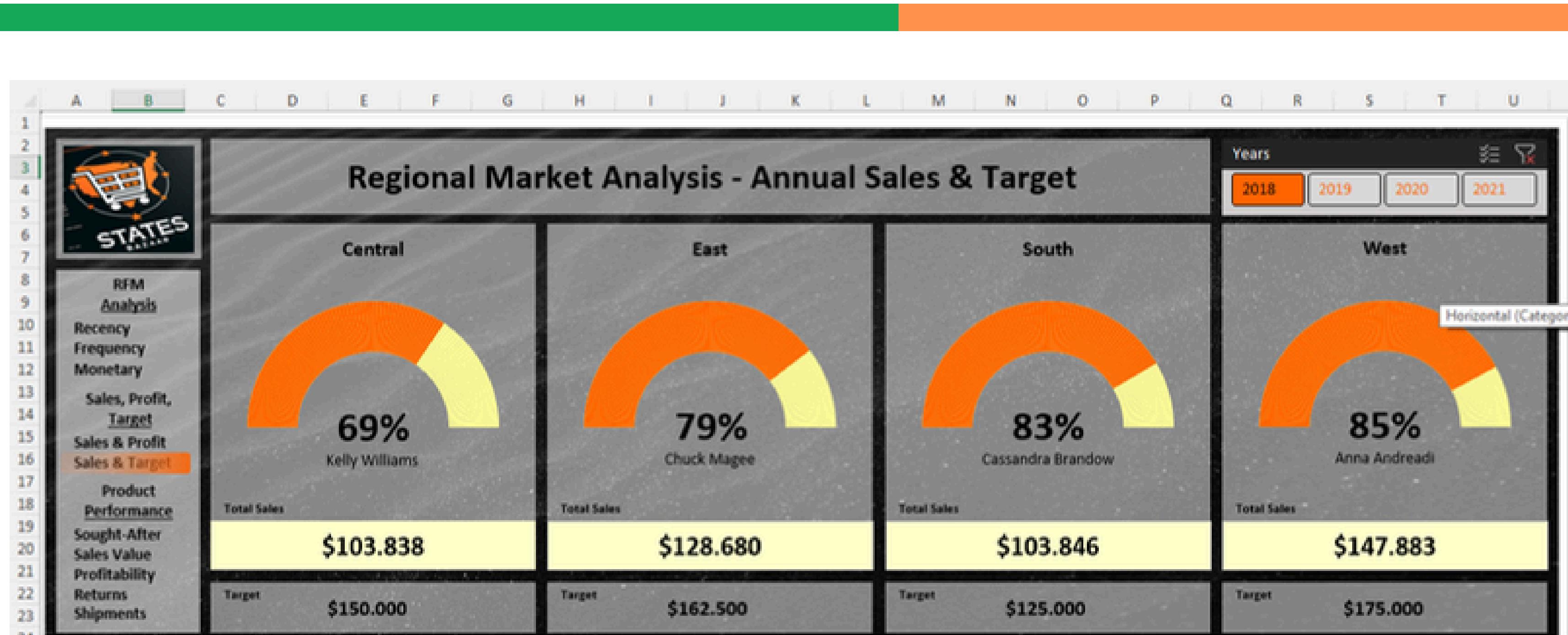
RFM Analysis - Monetary

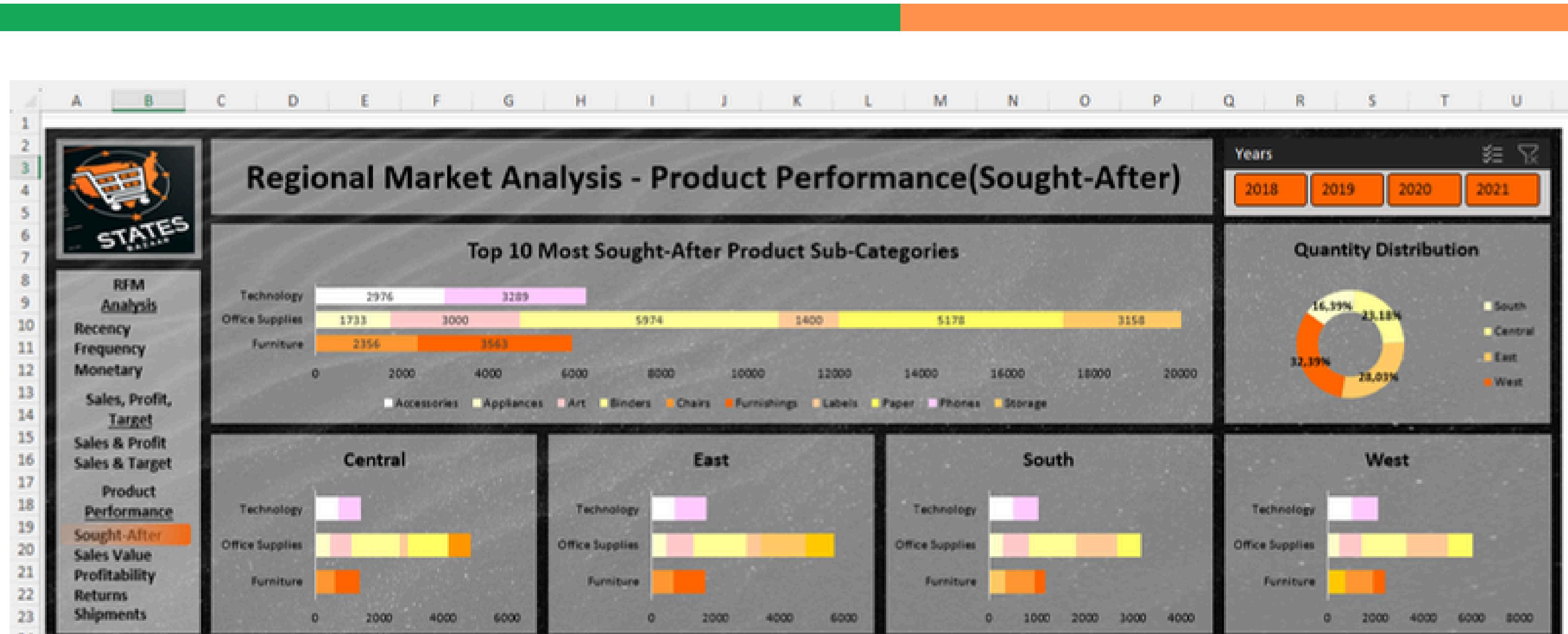


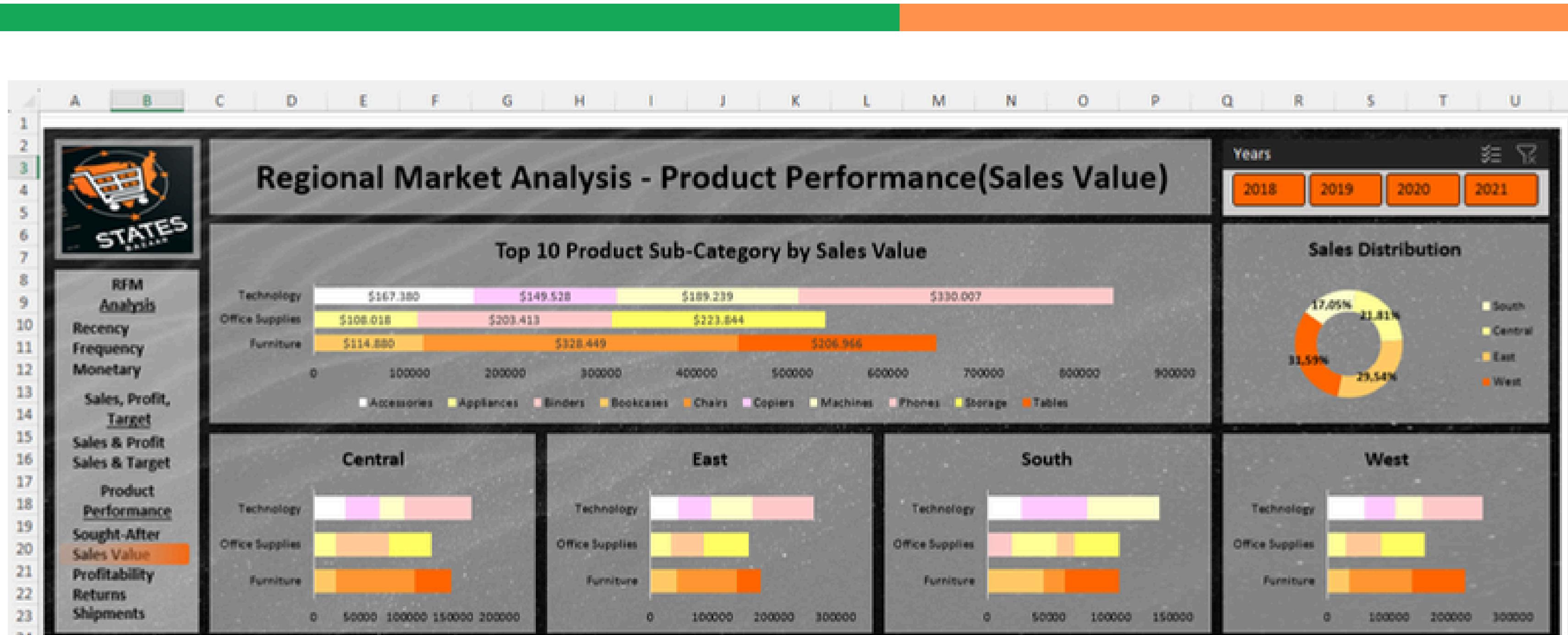


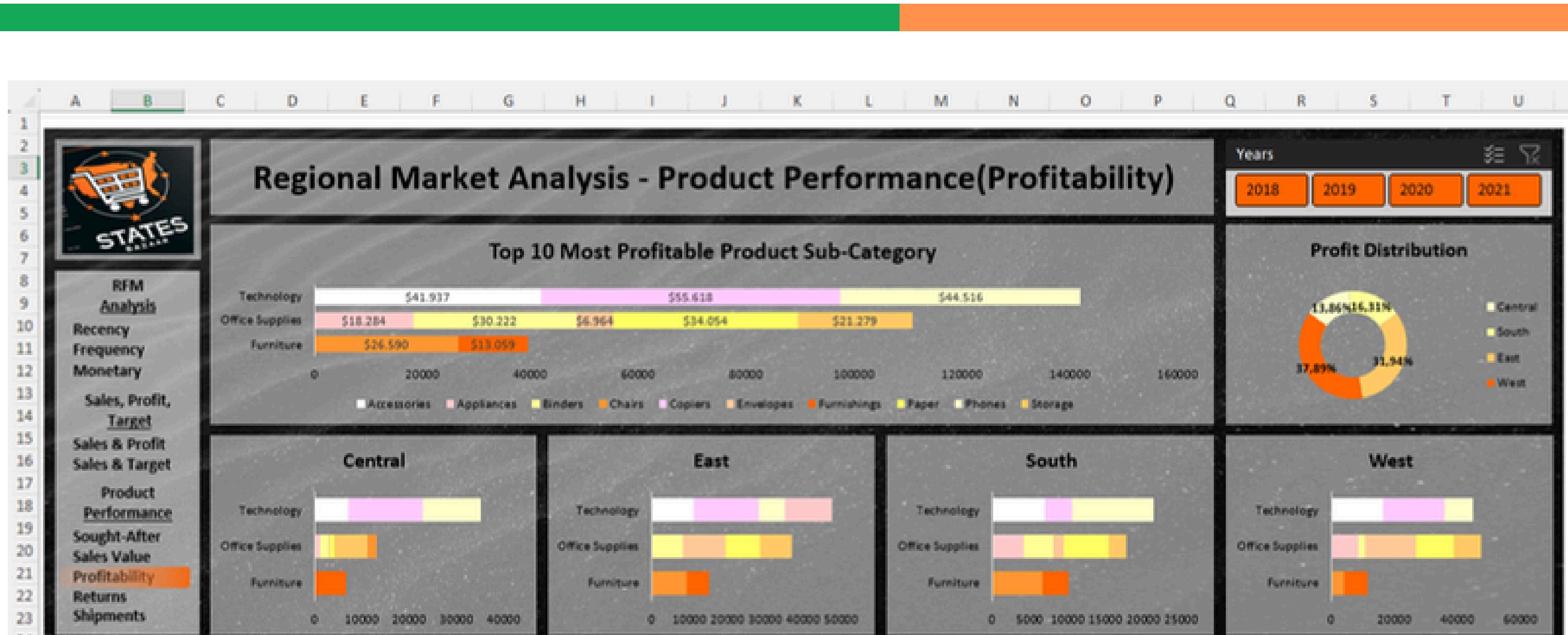
Frontend

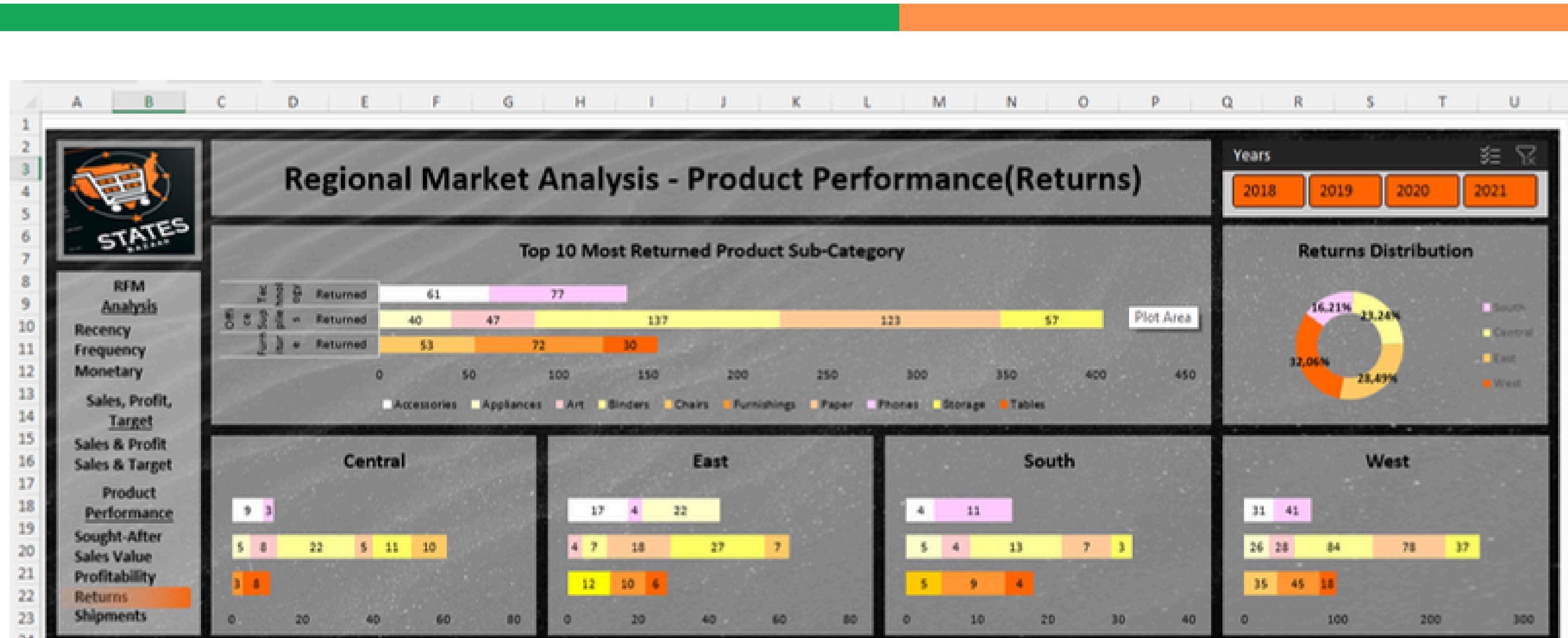
Sales, Profit, & Target Analysis

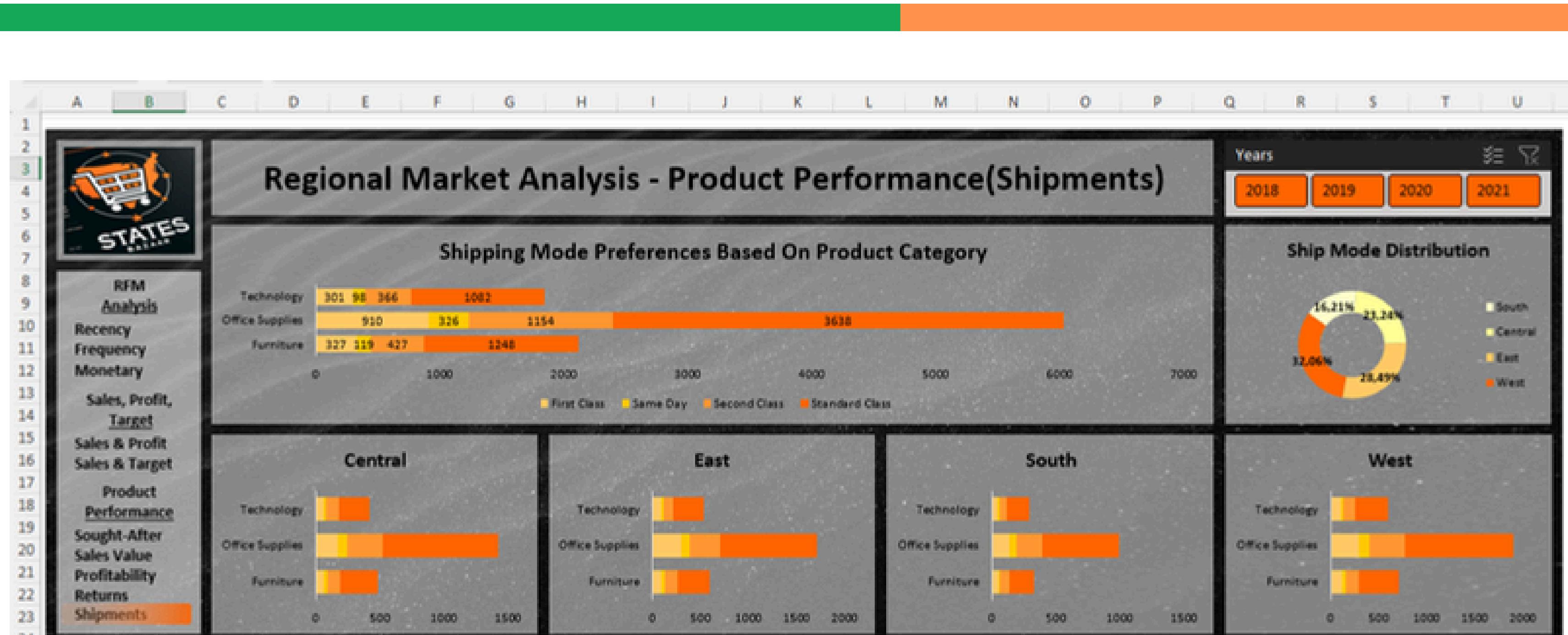














Analysis Results and Insights



In the following pages, I will present the key findings derived from the data visualization dashboard. These findings will be organized into three main areas: RFM Analysis, Sales-Profit-Target Analysis, and Product Performance Analysis.

Each finding will be translated into actionable insights to support data-driven decision-making, contributing to the company's strategic planning and long-term business growth.



RFM Analysis

Key Findings – Summary and Monthly Trends

Recency

General Insight:

The highest purchasing activity occurred in February and April, while January and September recorded the longest intervals between transactions, indicating lower recency.

Regional Highlights:

- Central: Peaks in March and October; lowest activity in August.
- East: Peaks in April and May; lowest in June and September.
- South: Peaks in April and June; lowest in January and December.
- West: Peaks in February and April; lowest in January and March.

Frequency

General Insight:

The highest transaction frequency was observed in November and December, while the lowest occurred in January and February.

Regional Highlights:

- A consistent pattern across all regions: frequency peaks at year-end and drops significantly at the beginning of the year.

Monetary

General Insight:

Sales volume peaked during November and December, while January and February saw the lowest sales.

Regional Highlights:

- All regions exhibit a strong year-end trend, reinforcing the fourth quarter as the primary shopping season.

RFM Analysis

Strategic Insights – Optimal Timing & Correlation

Year-End Shopping Season (November–December) as a Golden Window

There is a strong correlation between frequency and monetary metrics during November and December, indicating a seasonal shopping pattern. This surge is likely driven by major promotional events (e.g., 11.11 and 12.12) and year-end holidays.

→ This period represents a strategic opportunity to maximize sales and customer engagement.

Beginning of the Year (January–February) as a Low Season

All key metrics decline significantly in January and February. This trend may reflect post-holiday spending fatigue and consumer focus on new financial planning.

→ The downturn occurs across all regions, indicating the need for targeted interventions or incentive strategies to stimulate demand.

Diverse Regional Performance Patterns

- East & South: Demonstrated stable recency performance in Q2 (April–June), suggesting strong customer engagement momentum during this period.
- Central: Showed high volatility, with peak performance in March and October and a notable drop in August. This may point to inconsistencies in promotions or logistical execution.
- West: Consistently performed well in February and April in both recency and monetary terms, potentially indicating a loyal customer base with early-year purchasing behavior.

RFM Analysis

Strategic Recommendations

1. Boost Q1 Engagement (January–February)

- Launch early-year loyalty programs or flash sales across all regions to stimulate demand during the low season.
- Focus on retargeting customers who demonstrated low transaction frequency in the previous quarter, to re-activate dormant segments.

2. Optimize Peak Season (November–December)

- Prepare inventory and logistics well in advance (Q3) to ensure readiness for the year-end surge.
- Segment promotional strategies by region:
 - Central and West: Position premium product offerings.
 - South and East: Emphasize bundled deals or volume-based discounts to drive conversion.

3. Region-Specific Campaign Strategies

- East and South: Capitalize on high customer engagement in April–June with targeted upselling campaigns.
- Central: Reassess monthly promotional and logistics strategies to minimize sharp performance drops (e.g., in August).
- West: Leverage strong early-year performance with exclusive campaigns—this region may serve as a testing ground for new product launches.



Sales, Profit, and Target Analysis

Key Findings and Summary

1. Sales Distribution (%) – West & East Lead the Market

- West consistently contributes the highest share of total sales (30–34%) each year, making it the most dominant region in terms of revenue generation.
- East experienced a significant rise in 2019 (33.22%) and maintained strong performance throughout 2020–2021, indicating a positive growth trajectory.
- Central remained relatively stable (~20–24%) but showed no significant growth.
- South recorded the weakest sales contribution (less than 17%), with a sharp decline observed between 2019 and 2021.

2. Profit Distribution (%) – West as the Most Profitable Region

- West was the top contributor to overall profit annually, accounting for up to 47% in 2021, reflecting high operational efficiency.
- East also delivered strong and consistent profit performance, maintaining a share above 30% and peaking at 35.51% in 2021.
- Central and South underperformed in profit relative to their sales contribution—particularly Central, which contributed only 8% of total profit in 2021, despite holding 20% of the sales share.

3. Target Completion Rate – Sharp Growth in 2020–2021

- All regions recorded a significant increase in target achievement starting from 2020.
- East and West consistently exceeded their targets from 2020 onward (East: 131% in 2021, West: 143%).
- Central and South reached near-target performance in 2021, both achieving 98% completion rates.
- South had a notable underperformance in 2019 (only 57%), warranting further investigation into possible causes such as supply chain disruptions or weak local demand.

Sales, Profit, and Target Analysis

Strategic Insights

West – The Golden Region

West not only contributes the highest share of total sales but also consistently delivers the largest share of overall profit.

→ This indicates both strong market potential and high operational efficiency, making it a key region for continued investment and optimization.

East – A Fast-Growing and Profitable Market

East has demonstrated rapid growth in sales and has exceeded sales targets for the past two consecutive years.

→ This suggests East is a strategic candidate for expansion initiatives and product innovation testing.

Central – Consistent Volume, Weak Profitability

While Central maintains stable sales volume, its profit margins are notably low and declined sharply in 2021.

→ Possible factors include high operational costs, heavy discounting, or low selling prices, which merit a thorough cost structure review.

South – Underperforming but Showing Improvement

Despite lower sales and profit levels, South achieved 98% of its target in 2021, indicating a potential recovery and room for improvement in sales execution and market strategy.



Sales, Profit, and Target Analysis

Strategic Recommendations

1. Prioritize West & East for Scaling and Product Launches

Establish West and East as flagship regions for major campaigns and new product launches, leveraging their proven sales and profit performance.

Strengthen customer retention programs and loyalty initiatives to sustain and further enhance profitability.

2. Reevaluate Pricing and Cost Structure in Central

Conduct a thorough profit margin audit in Central to identify potential issues related to pricing models, logistical costs, or overly aggressive discounting.

Consider implementing cost-efficiency strategies or repositioning product offerings to align with regional profitability goals.

3. Redesign Market Strategy for South

Develop a more granular market segmentation approach—focus on niche segments or product categories that align with local preferences.

Invest in distribution channels, brand awareness, and customer education to establish a stronger foundation for sustainable growth.

4. Capitalize on 2020–2021 Performance Momentum

With all regions showing significant improvement in target achievement over the past two years, this is a timely opportunity to increase performance targets and adopt a more aggressive strategy moving forward.

5. Implement Data-Driven Forecasting for Target Setting

Annual target-setting should be dynamically adjusted based on regional performance trends rather than static assumptions.

For instance: increase targets for East and West, maintain a stable target for Central, and gradually raise expectations for South in line with its recent improvements.

Product Performance Analysis

Key Findings and Summary

1. Most Sought-After Products (Demand)

- Overall, the most in-demand products are Binders, Paper, and Phones.
- Binders and Paper consistently rank high in demand across all regions.
- West and South show additional demand for technology and creative products, such as Phones and Art supplies.
- Product quantity sold is dominated by West (32.39%) and East (28.03%), indicating the highest market demand concentrations.

2. Sales Value

- The top products by sales value are Chairs, Phones, and Storage units.
- Phones remain a high-revenue product across all regions, reflecting wide market penetration.
- Total sales continue to be dominated by West and East (combined 61.13%).

3. Profitability

- The most profitable product categories are Copiers, Phones, and Accessories.
- Copiers consistently lead in profitability across Central, East, and West – highlighting strong profit margins.
- South relies more on Phones and Accessories, which remain profitable, though with lower margins than Copiers.
- West (37.89%) and East (31.94%) jointly account for nearly 70% of total profit.

Product Performance Analysis

Key Findings and Summary

4. Returns

- The categories with the highest return rates are Binders, Paper, and Phones — the same categories that are most in demand.
- West (32.06%) and East (28.49%) have the highest return volumes, likely due to their high transaction volumes.
- Central and South exhibit lower return rates, with South being the lowest (16.21%).

5. Shipping Preferences by Product Category

- Office Supplies and Furniture are primarily shipped via Standard Class, reflecting cost-efficiency.
- Technology products (including Phones) still use Standard Class, but show slightly higher usage of First and Second Class options compared to other categories.
- Same-Day delivery is rarely chosen, indicating low urgency in product acquisition.

Product Performance Analysis

Strategic Insights

Demand ≠ Revenue ≠ Profit

- Products like Binders and Paper are highly sought-after but do not significantly contribute to revenue or profit and have high return rates.

Phones: A Strategic Triple Threat

- Phones are high in demand, revenue-generating, and relatively profitable – making them ideal for flagship campaigns.

Copiers: Low Volume, High Margin

- Despite not being widely purchased, Copiers consistently yield the highest margins – suggesting a niche premium product position.

Returns Concentrated in High-Demand Categories

- High return rates in Binders, Paper, and Phones may indicate issues with customer expectations, product quality, or unclear product descriptions.

Shipping Behavior Reflects Price Sensitivity

- Most consumers opt for Standard or Second-Class shipping, even for tech products – revealing a strong preference for cost-efficiency over speed.



Product Performance Analysis

Strategic Recommendations

1. Improve Management of High-Demand, High-Return Products

- Reassess the design, quality, and product information for Binders, Paper, and Phones.
- Highlight product reviews and ratings more prominently.
- Consider implementing basic warranties or limited return policies to curb unnecessary returns.

2. Optimize Phones as a Flagship Product

- Leverage Phones' unique position by using them as the central focus of marketing campaigns.
- Offer bundled promotions with Accessories to increase average transaction value.

3. Promote High-Margin Products Like Copiers

- Launch educational campaigns to highlight the value and features of Copiers, especially for businesses.
- Target B2B and small business (SME) segments with tailored promotions.

Product Performance Analysis

Strategic Recommendations

4. Align Shipping Strategy with Consumer Preferences

- Prioritize optimization of Standard and Second-Class shipping, including cost negotiations with logistics partners.
- Use premium shipping selectively, such as for priority customers or exclusive campaigns.

5. Develop Region-Specific Differentiation Strategies

- West & East: Focus on high-volume, high-margin products (Phones, Copiers, Accessories).
- Central: Reduce return rates and promote profitable categories.
- South: Focus on building demand and improving customer education to establish market reliability.

Links and Resources



<https://github.com/evanserlangga/Portfolio/tree/main/States%20Bazaar>



https://drive.google.com/drive/folders/1Gcff4CVfZRCfqHPx-QG3l-E_6cDd-0m7?usp=sharing

My other Portfolios:



<https://github.com/evanserlangga/Portfolio>



<https://drive.google.com/drive/folders/1x1dcO58XjpokD4KDw4I0xtwmtpckQgW3?usp=sharing>

A photograph showing several farmers in a rice field at sunset. They are silhouetted against the bright sky, wearing traditional hats and carrying large woven baskets on their heads. The field is filled with large sacks of rice, and the background shows a distant town under a cloudy sky.

THANK YOU

STATES BAZAAR

IGNATIUS EVANS ERLANGGA